

A Space Communications Study

Final Report

September 15, 1966 - September 15, 1967

Prepared for

National Aeronautics & Space Administration

Electronic Research Center

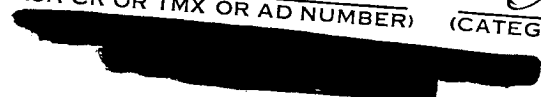
under

NASA GRANT NGR-33-006-020

GPO PRICE \$ _____
CFSTI PRICE(S) \$ _____
Hard copy (HC) 7.00
Microfiche (MF) 65

ff 653 July 65

DEPARTMENT OF ELECTRICAL ENGINEERING
POLYTECHNIC INSTITUTE OF BROOKLYN

FF No. 602(A)	<u>N68-10729</u> (ACCESSION NUMBER)	(THRU)
	<u>210</u> (PAGES)	<u>1</u> (CODE)
	<u>NR-90095</u> (NASA CR OR TMX OR AD NUMBER)	<u>07</u> (CATEGORY)
		

A Space Communications Study

Final Report

September 15, 1966 - September 15, 1967

Prepared for

National Aeronautics and Space Administration

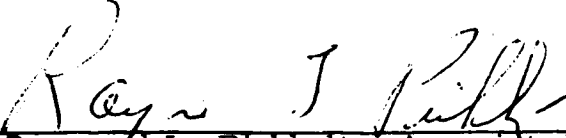
Electronic Research Center

under

NASA GRANT NGR-33-006-020

Principal Investigators


Kenneth K. Clarke, Professor


Raymond L. Pickholtz, Associate Professor


Donald L. Schilling, Associate Professor

DEPARTMENT OF ELECTRICAL ENGINEERING
POLYTECHNIC INSTITUTE OF BROOKLYN

TABLE OF CONTENTS

	Page
Introduction	
I. Threshold Extension	4
A. Maximum Likelihood Estimation	4
B. Phase Locked Loop Studies	20
1. Models	20
a. Model using most likely noise	22
b. Pulse Model for Input Noise	45
2. Fokker Planck Techniques	65
3. Drift Correction in VCO	89
C. The Frequency Locked Loop	100
II. Water Tank Channel Simulator	107
III. Partial Coherent Binary PSK Synchronization	121
IV. Recursive Detection	152
V. FM Devices	169
A. Wideband FM Generation	169
B. SNR for Bandpass Limiters	177
C. A Statistical Approach to Interference and Multipath in FM	191
VI. New and Continued Research	200
VII. Theses and Dissertations	204
VIII. Papers Published	206

Introduction

This final report summarizes all the research sponsored by the National Aeronautics and Space Administration under grant NGR-33-006-020 for the period Sept. 15, 1966 through Sept. 15, 1967. The research supported by this grant principally encompasses the problem of transmitting and receiving analog and digital signals through noisy media. A good deal of the study was spent in the analysis of FM signals corrupted by additive gaussian noise and received through a fading medium. Particular attention was focused on the problem of threshold extension. Throughout the study, theory and experiment were worked hand-in-hand. Approximately equal effort was spent on each.

Part I of this report presents the results of the threshold extension studies. The maximum likelihood estimation problem, for an analog FM signal lasting a fixed interval, is solved with the aid of a digital computer. In the area of sub-optimum receivers, two models characterizing the clicks are presented in conjunction with an analysis of the PLL. Each model is used in the calculation of the expected number of clicks for different 1st order PLL's and a comparison to the standard FMD. Studies performed using Fokker-Planck techniques are also shown. The output clicks of an FM receiver result in the occurrence of the threshold effect. Elimination or reduction of the effect of these clicks are studied by actual detection and cancellation of these clicks for an analog signal (performed in 1966 for digital signals), and also by using an FLL, a new parametric type receiver developed at PIB which inherently reduces the effect of clicks.

In Part II, the development of the PIB water tank channel simulator is discussed. The flexibility of this device has been greatly extended to permit a wider variety of experimental work requiring transmission of signals through various fading media.

Part III deals with the analysis of PSK digital signals transmitted through noisy media. Detection techniques, synchronization and error rates are presented.

Part IV shows the development of recursive detection, a different formulation of the classical problem of detecting signals in noise. The advantages of this method appear immediately to lie mostly in the area of digital signals. The characteristics of the transmission media permit calculation of the

likelihood function in terms of the most recent samples. This formulation also permits a direct determination of the performance of the optimum receiver under a broad class of noise conditions and is valuable in applications such as signal design and the study of the effect of the variation of system parameters.

Part V shows the development of a wideband FM generator, the analysis of limiters for FM signals in SNR and a statistical approach to multipath.

The results of the studies yield a good appraisal of the operation of the techniques and devices mentioned. New approaches to the various problems were suggested by many of the investigations and some of these have already been analyzed in this report. Others had to be left for continued and new research and are enumerated in Section VI.

This grant has also served to support the publication of a large number of papers, and Masters and Ph.D. theses listed in Sections VII and VIII.

The participants in this program were:

Professors R. Boorstyn
K. Clarke
P. Crepeau
D. Hess
R. Pickholtz
H. Schachter
D. Schilling
H. Taub

Messrs. A. Guida
E. Hoffman
E. Nelson
J. Oberst
P. Osborne
N. Tepedelenlioglu
M. Unkauf

The final report was prepared by:

Professors K.K. Clarke (I. B1b, I.D, II, V.A.)
R.L. Pickholtz (I.B.2, IV, V.C)
D.L. Schilling (I.A, I.B1a, I.C, III, V.B)

I Threshold Extension

I.A Maximum Likelihood Estimation*Introduction

The problem of designing optimum receivers for demodulating analog signals subject to additive gaussian noise has been handled satisfactorily in the case of linear modulation (AM), but the cases of nonlinear modulation (FM, PM) are still, for the most part, unsolved. Numerous articles have been written on the PLL and FMFB but none of the designs lead to devices which are optimum in some commonly accepted sense (least mean square error or minimum variance (MV), maximum likelihood (ML), etc.). Some may, in actuality, be very close to optimum devices, but this fact has not yet been demonstrated.

The ML estimator for an analog FM signal has been investigated, in order to extend the present knowledge of optimum FM reception (1). The primary objective is the solution of the ML equation to determine the FM threshold. In addition, the output signal-to-noise power ratio obtainable with the ML estimator at high input carrier-to-noise power ratios will be examined.

The first step toward solving the problem of designing an optimum PM receiver was made by Lehan and Parks (2) in 1953, when they considered the problem from the point of view of statistical estimation theory. Unfortunately, the mathematics presented in their paper is not clear enough for a complete understanding of their work, although the equations they derive appear to be correct. In 1954 Youla (3) placed their work on a firm mathematical basis. He derived a pair of nonlinear integral equations whose solution is the ML estimate of the modulation in the PM case. Subsequently

*

This report summarizes part of Technical Report T-3, "Optimal FM Demodulation" by Guida and Schilling. This work was sponsored by NASA-ERC under grant NGR -ee-006-020.

in 1964, Becker and Lawton (4) extended the results to the FM case, again reducing the problem to one of solving a pair of nonlinear equations. In 1964, Schwartz(5) demonstrated that the Youla equations could be derived using vector space concepts, which simplified the derivation considerably.

However, in none of these studies did the authors show how to solve the integral equations, other than to state that the equations closely resembled the phase-locked loop. In 1963, Van Trees (6) suggested an iterative technique for solving the equations in the white noise case where the two equations can be reduced to a single equation. However, this technique was found to perform successfully only far below threshold.

Van Trees' (6) simple iteration procedure for solving the ML equation was improved upon by employing an additional function called the likelihood function, which provides a measure of the likelihood of a test solution. The new procedure consists of generating a sequence of test solutions by varying the parameter ϵ in the following equation:

$$\text{Test Solution} = \text{Approximate Solution} + \epsilon \cdot [\text{Solution error due to the Approximate Solution}]$$

where the solution error is defined as the error produced by substituting the approximate solution in the ML equation. For each value of ϵ selected, the likelihood function of the associated test solution is evaluated; the test solution for which the likelihood function is a maximum replaces the approximate solution. The solution error and the integrated absolute value of the solution error (called the total solution error) are evaluated and the process is repeated. When the total solution error goes below some preset value, the process stops. Otherwise, the process is stopped after a preset number of iterations. The last test solution is then an approximate solution to the ML equation.

There are several advantages obtained with the new process:

1. The process can be used at any CNR. [The simple iteration procedure of Van Trees is limited to low CNR] ,
2. The process requires fewer iterations than the simple iteration procedure; however, the iterations in the new process are more complex,
3. The computer time required for this process is less than the simple iteration computer time, when the initial test solution is far from the final solution because this process can accept large test solution increments without diverging,
4. The likelihood associated with the solution found by this process must be greater than or equal to the likelihood of the solution found by the simple iteration process. It has not been demonstrated, however, that the solution obtained is the absolute maximum likelihood solution.

A computer simulation of the process for solving the ML equation provided evidence that the technique leads to a convergent series of approximate solutions . A number of sequences of signal and noise values were tested and the SNR-CNR results were plotted on a graph along with the asymptotic SNR-CNR curves. An attempt to measure the threshold of the ML estimator was unsuccessful because the amount of computer time required to perform this task was found to be excessive.

Derivation of the Maximum Likelihood Estimate

The modulation in a general communications system, during a time interval 0 to T, can be regarded as a particular sequence of voltages taken from the set of all possible sequences that could occur. From a knowledge of the modulation source, one can assign a probability of occurrence to each sequence. Similarly, the noise in the system, during the time interval 0 to T, can be regarded as a particular sequence of voltages taken from the set of all possible sequences that could occur. Again, knowing the properties of the noise source, one can assign a probability of occurrence to each noise sequence. At the receiver, in this communications system, one has a received voltage sequence consisting of a noise sequence plus a particular function of a modulation sequence. The ML estimate of the modulation sequence is then the most likely modulation sequence to have produced the received voltage sequence. In mathematical terms, the maximum likelihood estimate is that modulation sequence which maximizes the conditional probability of all possible modulation sequences given a particular received voltage sequence.

If $m(t)$ = modulation

Phase = $\int_0^t m(\lambda) d\lambda$ (for an FM system)

$S(t) = S(t, m)$ = Signal which is a function of the modulation -
AM, PM, FM

$n(t)$ = noise

$v(t)$ = received voltage

= $S + n = A \cos (\omega_c t + \int m) + n$ (for an FM system)

$a(t)$ = ML estimate of $m(t)$

$p(m|v)$ = condition probability of the modulation m , given
a received voltage, v

then

$a(t)$ = the solution of the equation

$$\left\{ \begin{array}{l} \frac{\partial}{\partial m} p(m|v) = 0 \\ \text{for which} \\ p(a|v) = \text{an absolute maximum} \end{array} \right. \quad (1)$$

$$(2)$$

If Eq. (1a) has more than one solution, then $a(t)$ must be found by evaluating $p(m|v)$ at each solution. Note also, that minimum likelihood estimates of m are also solutions to Eq. (1a).

For the case of an FM signal given by

$$S(t, m) = A \cos (\omega_c t + \int m) \quad (3)$$

the ML equation for FM can be put into two forms

Lawton's Form

$$a(t) = \frac{1}{N_o} \int_0^T \left[\int_0^u R_m(t, \lambda) d\lambda \right] [-A \sin(\omega_c u + \int a)] [A \cos (\omega_c u + \int m) + n - A \cos (\omega_c u + \int a)] du \quad (4)$$

Guida's Form

$$a(t) = \frac{1}{N_o} \int_0^T R_m(t, \lambda) \left[\int_{\lambda}^T [-A \sin(\omega_c \mu + \int a)] [A \cos(\omega_c \mu + \int m) + n - A \cos (\omega_c \mu + \int a)] d\mu \right] d\lambda \quad (5)$$

In these equations, the additive noise $n(t)$ is assumed to be "white" and normally distributed. If we write $n(t)$ as

$$n(t) = x(t) \cos \omega_c t - y(t) \sin \omega_c t \quad (6a)$$

and let

$$s(t) = \sin \int m + y \quad (6b)$$

and

$$c(t) = \cos \int m + x \quad (6c)$$

the ML Equation becomes

$$a(t) = \frac{A^2}{2N_o} \int_0^T \left[\int_0^u R_m(t-\lambda) d\lambda \right] (s \cos \int a - c \sin \int a) d\mu \quad (7)$$

Solution of the Maximum Likelihood Equation for an FM Signal

The basic technique used on the computer was to evaluate the function $a_R(t)$, given by

$$\begin{aligned} a_R(t) &= \frac{A^2}{2N_o} \int_0^T R_m(|t - \lambda|) \int_{\lambda}^T [s \cos \int a_i - c \sin \int a_i] d\mu \quad (8) \\ &= \text{result of substituting } a_i \text{ in the ML equation} \\ &= a_R(a_i) \end{aligned}$$

for a succession of $a_i(t)$ (estimates of the solution) in some systematic manner, so that

$$\int_0^T |a_R(\gamma) - a_i(\gamma)| d\gamma = \text{Total solution error (TSE)} \quad (9)$$

becomes arbitrarily small. Clearly, when $a_R(t) = a_i(t)$ the total solution error is zero and the equation is solved.

Several iteration techniques were attempted including that of Van Trees (6). These techniques all failed to converge. The technique employed, which did converge, used, in addition to the ML equation of Eq. 8, the likelihood function

$$L(a_i) = - \frac{1}{2} \left[[v-S]^T [R_n]^{-1} [v-S] + [m]^T [R_m]^{-1} [m] \right] \Bigg|_{m=a_i} \quad (10)$$

Then the following steps were followed

1. Let $a_o(t) = 0$
2. Using Eq. 8 evaluate $a_R(a_o, t)$
3. Let $a_1(t) = \epsilon_1 a_R(a_o, t)$
4. Vary ϵ_1 until
 $L(a_1) = \text{a maximum}$
The $a_1(t)$ obtained is called a_{im}
5. Let $a_2(t) = a_{im} + \epsilon_2 (a_R(a_{im}) - a_{im})$
6. Vary ϵ_2 until
 $L(a_2) = \text{a maximum}$
7. Continue the process until TSE (a_i) (Eq. 9) is arbitrarily small.

More than one maximum will often be found in step 4, indicating that more than one solution is possible. Since the absolute maximum likelihood estimate of m (Eq. 2) is being sought, and since $f(a_i)$ is a measure of the likelihood of a_i , the best a_i to choose among the available set of a_i appears to be the a_i corresponding to an absolute maximum of $f(a_i)$. It should be pointed out, however, that this does not guarantee that the final result has an absolute maximum likelihood. As far as is known, such a guarantee can only be obtained through a test of all possible vectors in the vector space, which is clearly impractical if not impossible.

An alternate procedure is to choose a_i so that the total solution error is a minimum. Since the maxima of the likelihood function and the minima of the total solution error do not coincide, the best one can do is to choose the minima of the solution error closest to the absolute maximum of the likelihood function. A test of this procedure indicated that it does not converge to the solution as fast as the first method.

Results

The ML equation was solved assuming that the modulation was a frequency pulse lasting a time T seconds (see Fig. 6) .

Fig. 1 shows a graph of the likelihood function, $L(\epsilon_1)$ obtained in accordance with step 4 of the procedure. Figs. 2 and 3 show the corresponding variations in the total solution error, $TSE(\epsilon_1)$ and the signal - to - noise ratio, $SNR(\epsilon_1)$.

$$SNR(\epsilon_1) = \frac{E(m^2)}{E(a_1 - m)^2} \quad (11)$$

Note that the $L(\epsilon_1)$ curve has several peaks, anyone of which could lead to a solution. However, it has only one absolute maximum. In addition, it is clear that the peak of $L(\epsilon_1)$ is a good choice because the $SNR(\epsilon_1)$ also indicated a peak, in the vicinity of the peak of $L(\epsilon_1)$. The significance of the fact that the peaks in the $SNR(\epsilon_1)$ and $L(\epsilon_1)$ coincide is very great; this is the first indication that demodulation is taking place, i. e., that the correct signal is being extracted from the noise, (Of course, the $SNR(\epsilon_1)$ curve is only possible when the modulation is known beforehand. In a realistic demodulation problem, this curve would not be available). The curves of $L(\epsilon_i)$, $TSE(\epsilon_i)$ and $SNR(\epsilon_i)$ are similar to those of $L(\epsilon_1)$, $TSE(\epsilon_1)$ and $SNR(\epsilon_1)$ respectively and will not be given.

Figures 4 and 5 show the plots of SNR , TSE , L and ϵ as a function of the iteration number. Note that the likelihood function L is a continuously increasing function (approaching some asymptotic value) . This must be so since at any iteration the choice $\epsilon = 0$, which keeps L constant, is available. However, if the best choice of ϵ for maximum L is $\epsilon = 0$, the iteration process would stop because all the values attained by the various functions at $\epsilon = 0$ would be repeated at all future iterations. If this situation had occurred, we would have used a constant for a_R in the next iteration and then returned to the normal routine.

Fig. 6 shows the approximate solution of the ML equation for the conditions stated earlier after 10 iterations. The large size of the TSE should not be regarded as an indication that the approximate solution is far from the true solution. For example, consider the equation

$$x = 1000 \sin (x-1) \quad (12)$$

for which

$$\text{TSE} = |1000 \sin (x-1) - x| \quad (13)$$

The correct solution is approximately $x = 1.001$; however, if $x = 1.1$ (which represents a 10% error) is substituted in Eq. 13, the TSE is 99.9, which is quite large. Thus, the TSE should be compared to the maximum value it can attain, which is approximately 1000; then the TSE of the 99.9 indicated a relatively good solution. The approximate maximum value of the TSE for the ML equation has been derived in appendix D of Ref. 1.

The result is

$$\text{TSE}_{\max} \approx 535 \quad (14)$$

The total solution error should be compared to this value for purpose of estimating how far away the approximate solution is from the true solution.

References

1. Optimum FM Demodulation, A. Guida and D. L. Schilling, Technical Report T-3 1963, Polytechnic Institute of Brooklyn, Electrical Engineering Department.
2. Optimum Demodulation
F. Lehan, R. Parks IRE Conventional Record IT, March, 1953.
3. The use of the method of maximum likelihood in estimating continuous - modulated intelligence which has been corrupted by noise.
D. Youla IRE Trans. Information Theory, March, 1954.
4. Results of Investigations of analog and digital communication techniques
H. Becker, J. Lawton, A. Kobos, IEEE Conventional Record Part 6
March, 1964.
5. Maximum a posteriori demodulation of analogue - type signals through random fading media
M. Schwartz Research Report PIBMRI - 1244-64, December, 1964.
6. The structure of efficient demodulators for multidimensional phase modulated signals
H. Van Trees IEEE Trans. Communication Systems. September, 1963.

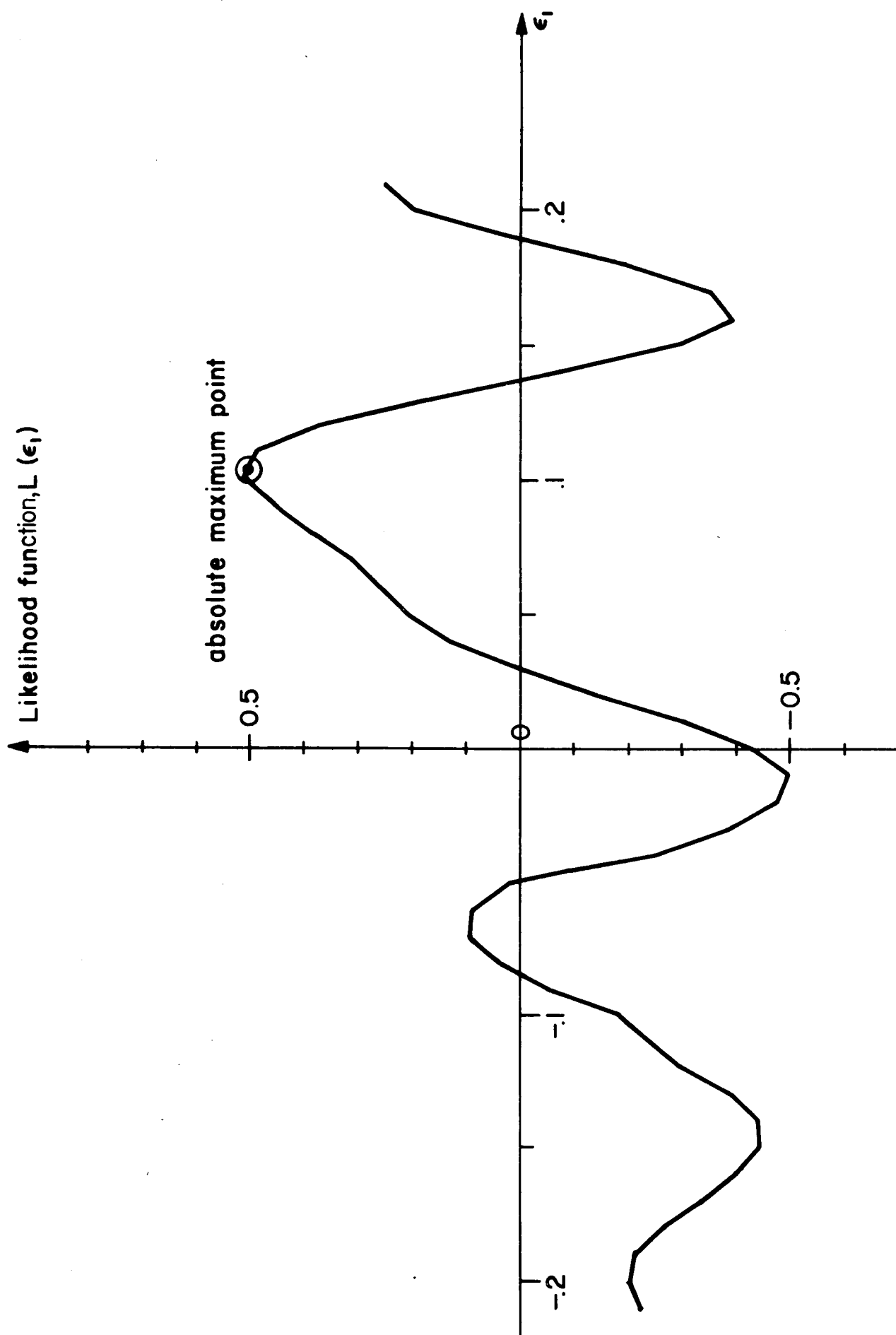


Fig. 1 Graph of the likelihood function versus ϵ_1

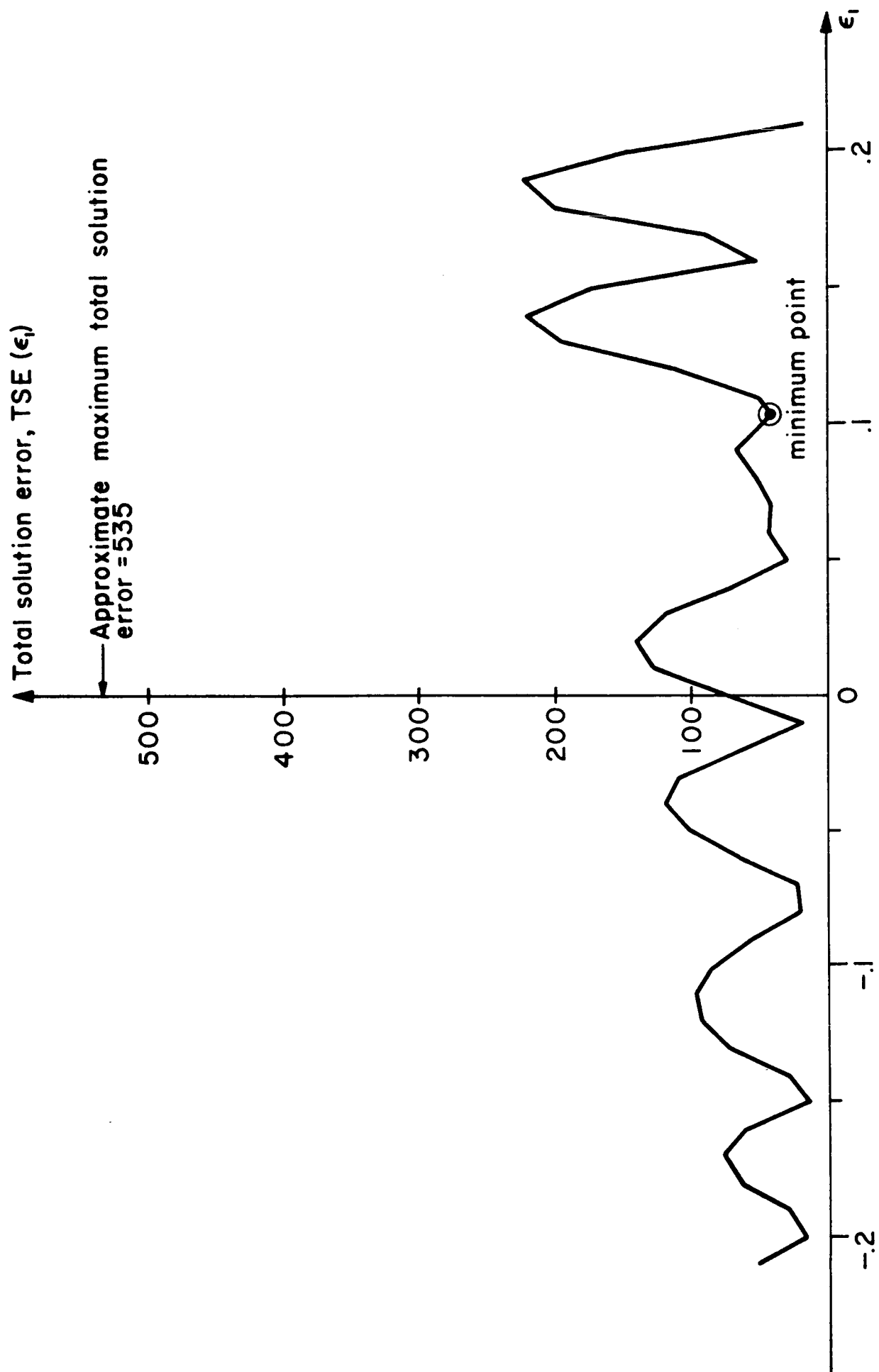


Fig. 2 Graph of the total solution error versus ϵ_1

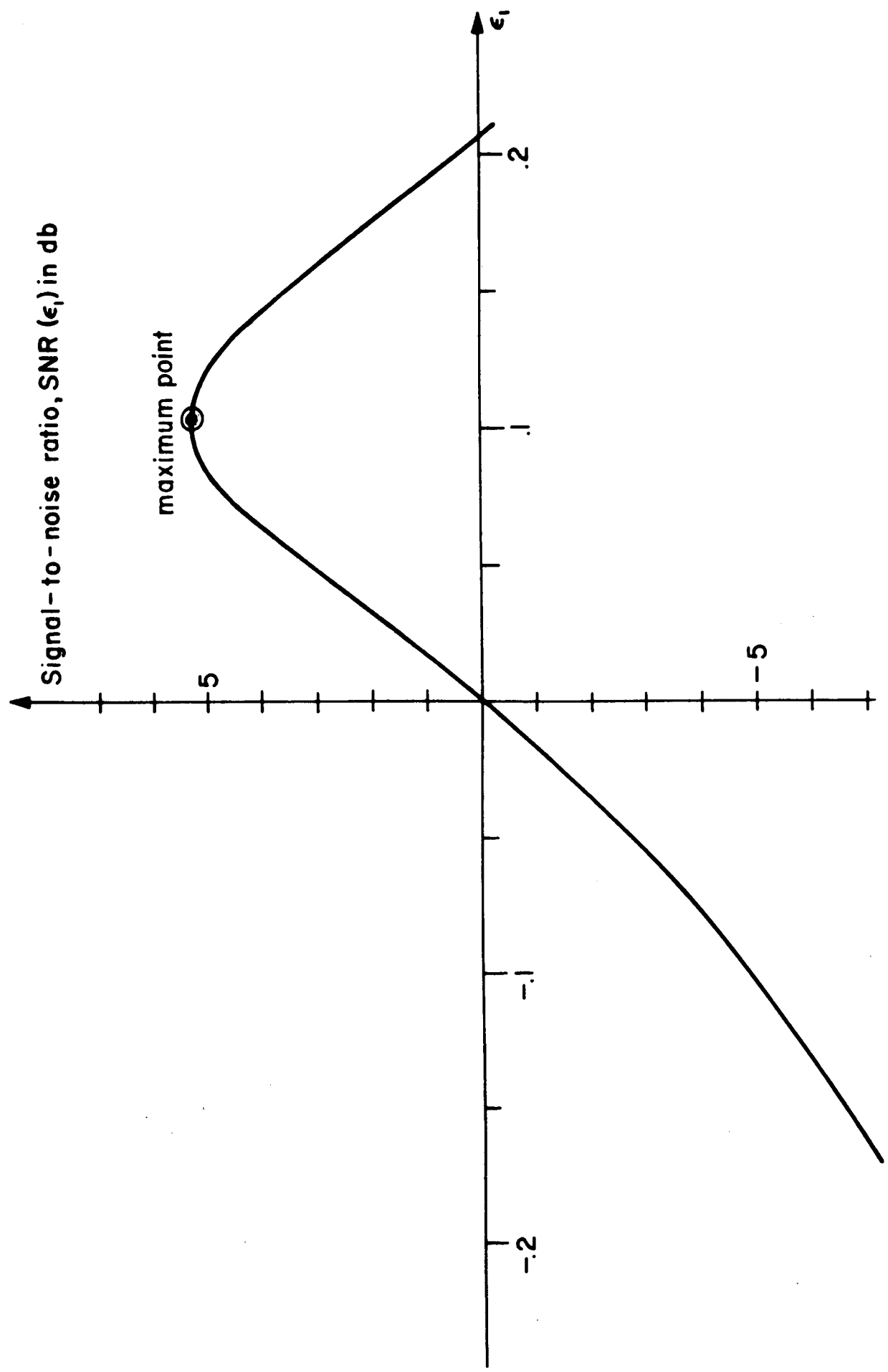


Fig. 3 Graph of the signal-to-noise ratio versus ϵ_1

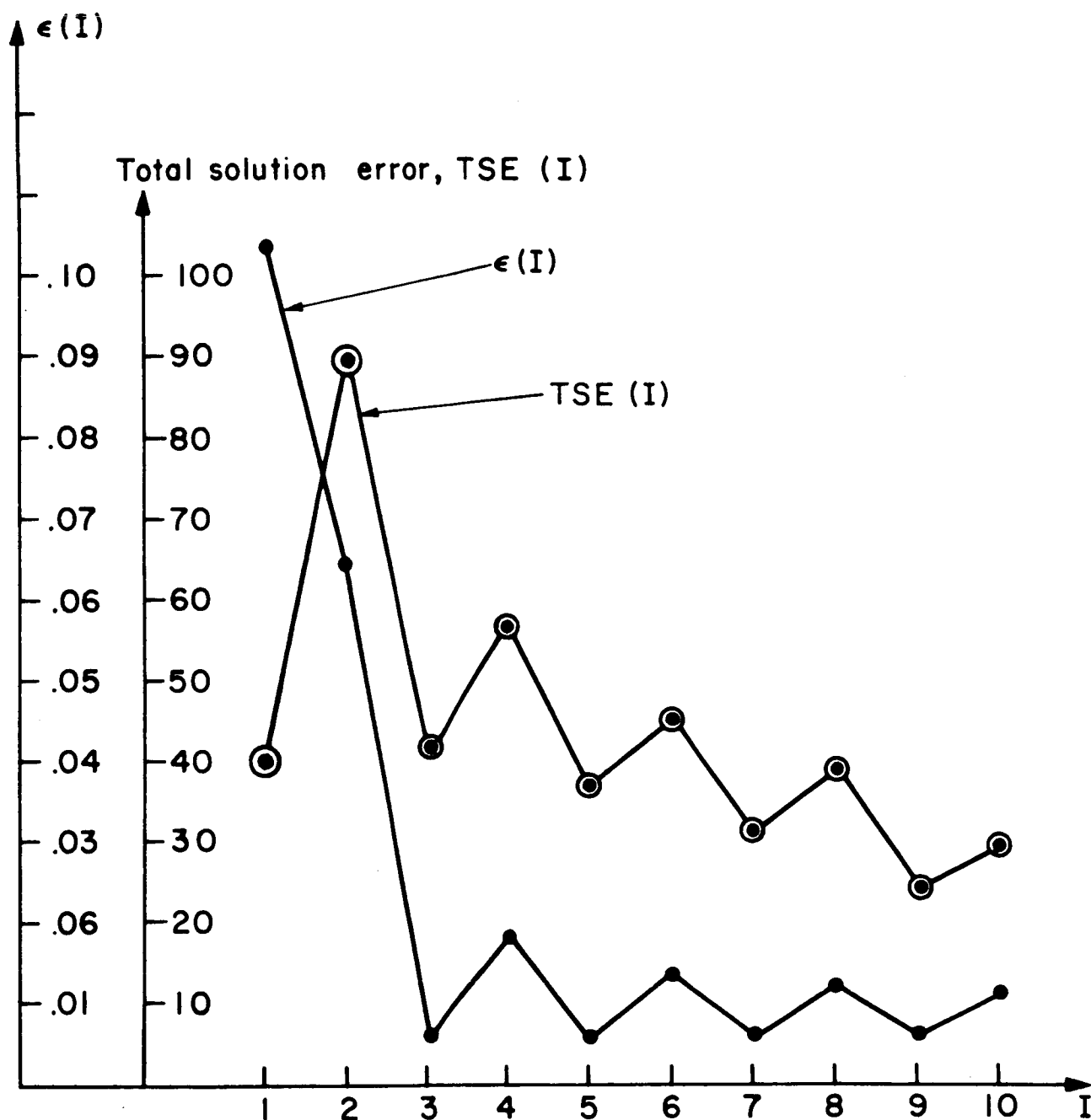


Fig. 4 Graph of ϵ and the total solution error versus the iteration number, I.

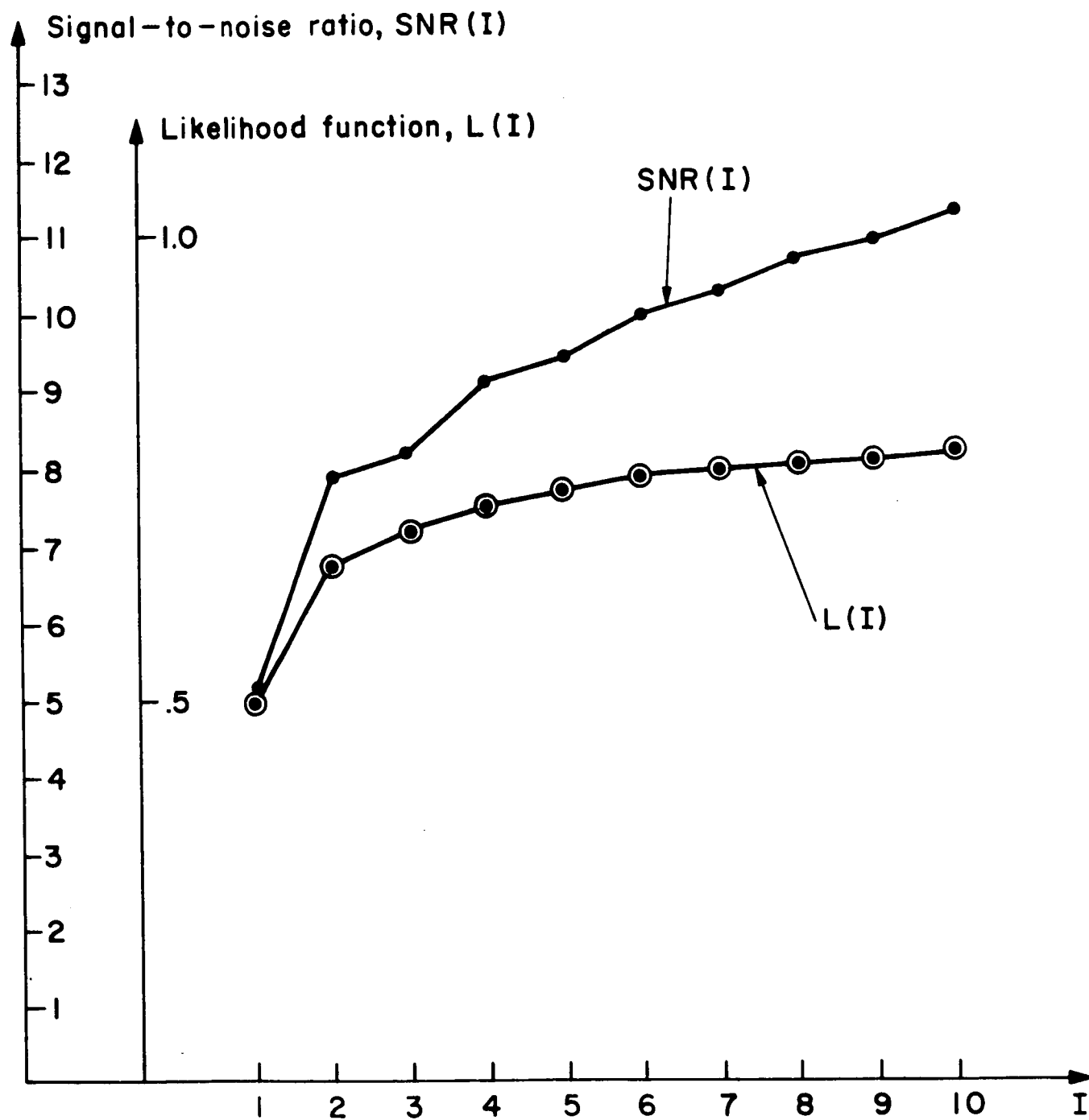


Fig. 5 Graph of the signal-to-noise ratio and the likelihood function versus the iteration number, I .

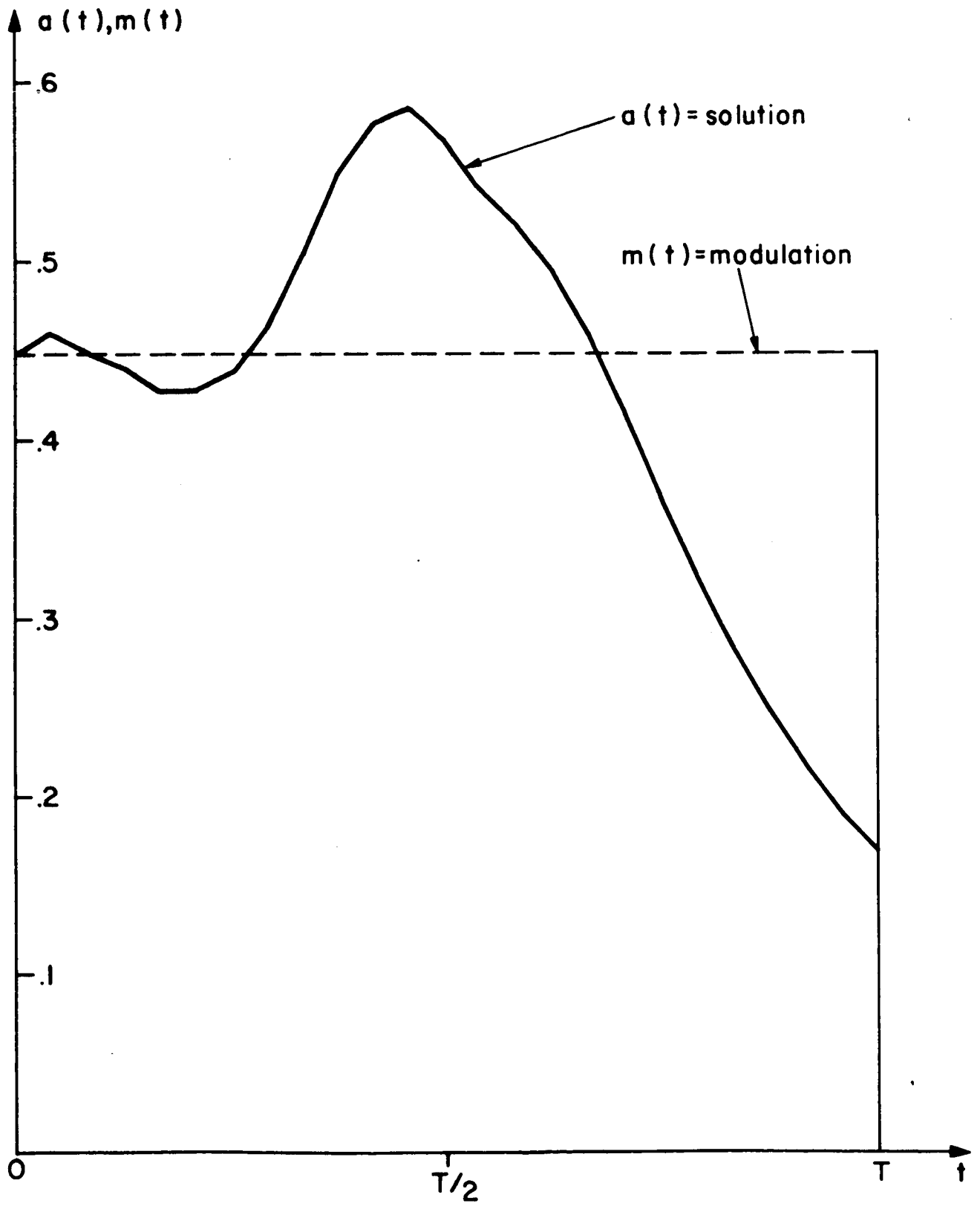


Fig. 6 Graph of the solution of the ML equation after 10 iterations.

I-B Phase Locked Loop Studies

I-B. 1 ModelsIntroduction

The response of the phase locked loop (PLL) to an FM signal in the presence of noise is describable as the solution of a time-varying, non linear differential equation driven by a random process. Since there is, at present no general method for extracting the desired response statistics by purely analytic methods, it has become necessary to devise appropriate and approximate models for the purpose of analyzing the performance of the PLL.

One of the most successful models, especially as regards the threshold performance of a PLL, is that based on the observation that the output of the PLL is experimentally observed to consist of 1) the demodulated signal 2) "smooth" or Gaussian noise and 3) "clicks". These clicks are relatively large pulses having a fixed net area corresponding to a total change in r.f. phase of 2π radians. Schilling has employed this technique to predict threshold for a phase locked loop using a method for calculating the average number of clicks per second based on previous work of Rice on FM discriminators.

Since it is generally agreed that the average number of clicks per second is one of the most important characteristics of the PLL, it is important to develop reliable methods for making these calculations. While some analytic results are available by the use of Fokker-Planck techniques they are for the case of an unmodulated carrier, white noise, and a first order loop.

In order to provide a reliable technique for predicting the performance of practical loops of higher order, with inputs consisting of modulated carrier plus non-white noise, several promising models for the click-generating mechanism have been developed. In this section of this final report, two such models are discussed. In section I B 1a, a method based on calculations involving

modelling the input noise by its most likely trajectory is discussed. In section I B 1b the model consists of approximating the input signal by a uniform envelope pulse and a Gaussian rate of change of phase (frequency). Numerical calculations have been carried out for both models. The first model has been simulated on a digital computer. Experimental verification has been obtained over the range of validity of both models.

I-B.1a Model Using Most Likely Noise

Introduction

This report presents a new method for estimating the expected number of clicks occurring per second at the output of an FM demodulator. The method is used here to determine the expected of number of clicks per second occurring at the output of a first order phase locked loop when a carrier signal embedded in gaussian noise is received. The technique can also be used to determine the response of higher order PLL's, with and without modulation, the Frequency Demodulator Using Feedback, and the Maximum Likelihood Estimator. This will be discussed in subsequent reports.

The phase locked loop considered here is shown in Fig. 1. The inphase component, $y(t)$, and the quadrature component $x(t)$, of the input noise (see Fig. 2), are each represented by the conditional expectation $y_1(t)$ and $x_1(t)$,

$$x_1(t) = E [x(t) / x(o), \dot{x}(o)] \quad (1)$$

$$y_1(t) = E [y(t) / y(o), \dot{y}(o)] \quad (2)$$

The time $t = 0$, is defined as the time when the quadrature noise component passes through zero, as shown in Fig. 2. Thus

$$x(o) = 0 \quad (3)$$

the "noise" terms $x(t)$ and $y(t)$ are deterministic functions depending on the initial conditions which will be used as parameters in our procedure, and on the impulsive response of the IF filter.

Varying the parameters $\dot{x}(o)$, $y(o)$, and $\dot{y}(o)$ determine a surface in $\dot{x}(o)$, $y(o)$, and $\dot{y}(o)$ space, such that on one side of the surface a click results, while on the other side of the surface no click results. (1). Knowing the probability density of these parameters and the equation of the surface, we then calculate, following Rice (2) the expected number of

clicks occurring per second.

Expected Noise The IF filter is assumed to consist of the cascade of two identical stages, each single tuned with a 3db bandwidth of two radians per second. The two stages together, therefore, have a 6db bandwidth of two radians per second. Thus, the low pass quadrature noise components have the spectrum

$$S(\omega) = S_x(\omega) = S_y(\omega) = \frac{4\sigma^2}{(\omega^2 + 1)^2} \quad (4)$$

where σ^2 is the variance of the random processes $x(t)$, $y(t)$.

It then follows that:

$$R(t) = R_x(t) = R_y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega t} d\omega = \sigma^2 (1 + |t|) e^{-|t|} \quad (5)$$

$$\begin{aligned} R_{xx}(t) = R_{yy}(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (-j\omega) S(\omega) e^{j\omega t} d\omega = -\frac{d}{dt} R(t) = \sigma^2 t e^{-|t|} \\ &= -\dot{R}(t) \end{aligned} \quad (6)$$

The conditional density of $x(t)$ given $x(o)$, $\dot{x}(o)$ is,

$$f(x(t)/x(o), \dot{x}(o)) = \frac{1}{\sqrt{2\pi\sigma^2 \left(1 - \frac{R^2(t)}{\sigma^4} - \frac{\dot{R}^2(t)}{\sigma^4}\right)}} e^{-\frac{\left(x(t) - x(o)\frac{R(t)}{\sigma^2} + \dot{x}(o)\frac{\dot{R}(t)}{\sigma^2}\right)^2}{2\sigma^2 \left(1 - \frac{R^2(t)}{\sigma^4} - \frac{\dot{R}^2(t)}{\sigma^4}\right)}} \quad (7)$$

Similarly $f(y(t) / y(o), \dot{y}(o))$ has the same form as $x(t)$. We can see by inspection that (using 5 and 6 for $R(t)$ and $\dot{R}(t)$):

$$x_1(t) = E[x(t)/x(o)=0, \dot{x}(o)] = \dot{x}(o) t e^{-|t|} \quad (8)$$

and

$$y_1(t) = E[y(t)/y(o), \dot{y}(o)] = \left\{ y(o) (1 + |t|) + \dot{y}(o)t \right\} e^{-|t|} \quad (9)$$

It should be noted that when using a double stage filter, the only initial conditions of importance are $x(o)$, $\dot{x}(o)$, $y(o)$ and $\dot{y}(o)$. Higher order filter require more derivatures and hence result in an increased number of parameters. These in turn result in a more complex surface. The locus of the envelope of the signal and noise when using Eqs. 8 and 9, and $\dot{y}(o) = 0$ in Fig. 3. Note that if $y(o) < 1$, there is no input phase jump.

PLL Differential Equation

The differential equation for the first order PLL having again G , and receiving an unmodulated carrier (see Fig. 1) is,

$$\ddot{\phi} + G \sin \phi = G(x_1(t) \cos \phi + y_1(t) \sin \phi) \quad (10)$$

where $x_1(t)$ and $y_1(t)$ are given by Eqs. 8 and 9.

Equation 10 is solved in the time interval

$$-5 \leq t \leq 4$$

using a digital computer. A Runge-Kutta procedure is used to start the solution, and Moulton's Predictor-Corrector method used to continue the solution (both of fourth order). The solution is restarted by the Runge-Kutta Method at $t = 0$ since Eqs. 8 and 9 have discontinuous derivatives at that point.

The $x(o)$, $y(o)$, $\dot{y}(o)$ Surface

For a selection of $y(o)$, $\dot{y}(o)$, a hunting procedure is used on $\dot{x}(o)$ to determine the boundary $x_b(o) = f(y(o), \dot{y}(o))$. If $\dot{x}(o) > \dot{x}_b(o)$ a click (phase jump) is obtained. If $\dot{x}(o) < \dot{x}_b(o)$ for a given $y(o)$, $\dot{y}(o)$ a click is not obtained. The surfaces indicated by the families of curves shown in Figs. 4, 5 and 6 show the nature of the boundary for loop gains of 1, 3 and 5. When the vector $\dot{x}(o)$, $\dot{y}(o)$, $y(o)$ lies above this surface, a negative click or a phase jump of -2π is obtained. There is a mirror image of

this surface not shown, on the other side of the $\dot{y}(o)$, $y(o)$ plane, such that if the vector $\dot{x}(o)$, $\dot{y}(o)$, $y(o)$ is below this image surface a phase jump of $+2\pi$ is obtained. The reason this mirror image surface exists is that the phase locked loop is symmetrical when balanced, yielding equal numbers of positive and negative clicks.

It is also worth noting that the surface is symmetrical with respect to $\pm \dot{y}(o)$. The reason for this is that the effect of changing $\dot{y}(o)$ to $-\dot{y}(o)$ is identical to changing t to $-t$ in the differential equation for the loop.

It should be pointed out that no surface could be found for $y(o)$ less than 1. This suggests that a phase jump can only be obtained in a PLL when a phase jump occurs in the input. (Figure 3 shows that when $\dot{x}(o) > 0$ and $y(o) > 1$, a phase jump of -2π radians occurs in the input, i.e. one complete negative rotation about 1.

The Expected Number of Clicks/second, N

To simplify the calculation of the probability of being above the surfaces shown in Figs. 4, 5, and 6, the curved surfaces (which are approximately arcs of circles) are approximated by the straight lines shown in the figures (a more exact determination of N obtained by using the arc of a circle is currently in progress). These curves in Figs. 7, 8 and 9 are bounded by straight lines as shown.

The probability of a negative click occurring in a time Δt is:

$$\begin{aligned}
 P(\text{negative click}) &= \sum_{i=1}^3 \text{Prob} \left(x(o) = 0, \dot{x}(o) > A_i y + B_i \text{ for } y_{1i} < y(o) < y_{2i} \right) \\
 &= \sum_{i=1}^3 \frac{\Delta t}{(2\pi\sigma^2)^{3/2}} \int_{y_{1i}}^{y_{2i}} \int_{A_i y + B_i}^{\infty} \dot{x} e^{-\dot{x}^2/2\sigma^2} e^{-y^2/2\sigma^2} dx dy
 \end{aligned} \tag{11a}$$

where $\dot{x}(o) = A_i y + B_i$ is the i^{th} straight line which extends from $y(o) = y_{1i}$ to $y(o) = y_{2i}$.

Note that the variance for $x(o) \dot{x}(o)$, $y(o) \dot{y}(o)$ are each equal to σ^2 .

The expected number of clicks per second is just $2P$ (negative click)/ Δt :

$$E[\text{click/sec.}] = N = \sum_{i=1}^3 \frac{2}{(2\pi\sigma^2)^{3/2}} \int_{y_{1i}}^{y_{2i}} \int_{A_i y + B}^{\infty} \dot{x} e^{-\dot{x}^2/2\sigma^2} e^{-y^2/2\sigma^2} dy \quad (11b)$$

Carrying out the integrations we get:

$$N = \sum_{i=1}^3 \left\{ \text{erf} \left[\left(\sqrt{A_i^2 + 1} y_{2i} + \frac{A_i B_i}{\sqrt{A_i^2 + 1}} \right) \frac{1}{2\pi\sqrt{A_i^2 + 1}} e^{-\left[\frac{B_i^2}{A_i^2 + 1} + \frac{1}{2\sigma^2} \right] \frac{1}{\sqrt{2\sigma^2}}} \right] \right. \\ \left. - \text{erf} \left[\left(\sqrt{A_i^2 + 1} y_{1i} + \frac{A_i B_i}{\sqrt{A_i^2 + 1}} \right) \frac{1}{\sqrt{2\sigma^2}} \right] \right\} \quad (11c)$$

The ratio $\frac{1}{2\sigma^2}$ can be recognized as the input carrier-to-noise ratio (CNR) since the amplitude of the carrier is taken to be unity.

Curves of the expected number of clicks per second, N , versus carrier to noise ratio are plotted in Fig. 10 for the gain 1, 3, and 5, for the upper and lower bounds obtained from the approximations to the curves of Figs. 7, 8, and 9.

Normalization

The results of Fig. 10, can be easily scaled to a PLL with a different 3dB bandwidth. Thus, the number of clicks/second when the 3db IF bandwidth is B_{IF} is:

$$N_{B_{IF}} = \frac{2\pi B_{IF} N}{2\sqrt{\sqrt{2}-1}} \doteq 4.881 B_{IF} N \quad (12a)$$

where $N_{B_{IF}}$ is the total number of clicks/second in a PLL with a 3dB IF bandwidth B_{IF} . N is found for a specified CNR and gain C from Fig. 10. (This G is obtained from Eq. 15 below). The gain G , is proportional to the ratio of loop to IF bandwidth. Thus,

$$G = 2 \left(\sqrt{\sqrt{2}-1} \right) \frac{B_L}{B_{IF}} \doteq 1.287 \frac{B_L}{B_{IF}} \quad (12b)$$

where

B_L is the 3db loop bandwidth.

The Effect of Some Other Deterministic Representations of Noise

The type of function used to represent the noise greatly affects the results obtained. Consider, for example, that the signal and noise at the phase detector input are represented by

$$v_i(t) = \begin{cases} \cos(\omega_0 t + \alpha t), & 0 \leq t \leq T \\ \cos \omega_0 t, & t \geq T \end{cases} \quad (13a)$$

where

$$\alpha T = 2\pi \quad (13b)$$

Then the PLL equation becomes

$$\begin{aligned} \dot{\varphi} + G \sin(\varphi - \alpha t) &= 0 & 0 \leq t \leq T \\ \dot{\varphi} + G \sin \varphi &= 0 & t \geq T \end{aligned} \quad (14)$$

Letting

$$\psi = \varphi - \alpha t \quad (15a)$$

yields

$$\dot{\psi} + G \sin \psi = -\alpha U(T - t) \quad (15b)$$

This equation is easily solved using standard integration techniques. The result of this integration shows that if

$$G > 0.675 \alpha \quad (16a)$$

a 2π rotation of the input signal and noise phasor ($\alpha T = 2\pi$), results in a click at the PLL output, and

$$|\psi| = |\varphi - \alpha t| < \pi$$

If

$$G < 0.675 \alpha \quad (16b)$$

the 2π radian rotation of the input signal and noise phasor, results in a doublet and ψ goes from 0 to 2π radians.

These results are shown in Figs. 11 and 12 for the case of α normalized to 1 rps.

When the "Most Probable" noise described by Eqs. 4 and 5 are employed, and $\dot{x}(o)$, $y(o)$ and $\dot{y}(o)$ adjusted so that

$$\sqrt{[1 - y(o)]^2 + [x(o)]^2} = 1 \quad (17a)$$

and

$$\frac{d}{dt} \tan^{-1} \left[\frac{x(t)}{1 - y(t)} \right]_{t=0} = \alpha \quad (17b)$$

i. e.,

$$x(o) = \dot{y}(o) = 0 \quad (17c)$$

$$y(o) = 2 \quad (17d)$$

and

$$\dot{x}(o) = \alpha = 1 \quad (17e)$$

then a gain of 0.725 is found to result in a doublet while a gain of 0.75 results in a click. These results are shown in Figs. 13 and 14.

If the most probable noise is perturbed slightly, so that

$$x(t) = \dot{x}(o) t e^{-1.25|t|} \quad (18a)$$

instead of

$$x(t) = \dot{x}(o) t e^{-|t|} \quad (18b)$$

where

$$\dot{x}(o) = \alpha = 1 \quad (18c)$$

and

$$y(t) = 2(1 + |t|) e^{-|t|} \quad (18d)$$

(see Eqs. 4, 5, 17c, 17d and 17e), then a gain of 0.90 results in a doublet, while a gain of 0.95 results in a click. These results are shown in Fig. 15'.

It is clearly seen that the shape of the noise model, as well as the initial conditions are of importance in determining the number of spikes occurring per second. In addition, since the shape of the noise depends on the impulsive response of the IF filter the noise model must be a function of the IF filter employed.

The Effect of Input Doublets

A careful investigation was made to see if a phase jump in the PLL could be obtained for noise $y(o) < 1$. In this case the input is a doublet instead of a phase jump of $\pm 2\pi$. Figure 16 shows the output of the loop with $\dot{x}(o) = 50$, $y(o) = 0.7$, $\dot{y}(o) = 0$. A doublet results at the PLL output. The large momentary gain of the loop due to a large $\dot{x}(o)$, results in a very large slope near $t = 0$. However, no spike was observed when the input was a doublet.

Threshold

If a rectangular low pass filter of bandwidth f_m , is placed at the output of the PLL, the output signal to noise ratio is:

$$\text{SNR} = \frac{\frac{\pi\beta^2 \text{CNR}}{2 F(\beta+1)}}{1 + \frac{4\pi N \text{CNR}}{2 F(\beta+1) f_m}} \quad (23)$$

where

$$F(\beta+1) = (\beta+1)^2 \left\{ \arctan\left(\frac{1}{\beta+1}\right) - \frac{\beta+1}{(\beta+1)^2+1} \right\}$$

$$f_m = \frac{1}{2\pi(\beta+1)} \quad (\text{base band})$$

and

N is the number of clicks occurring per second.

If the threshold is defined on that point where the denominator of Eq. 19 is 1.25,

$$\frac{4\pi N \text{CNR}}{2 F(\beta+1) f_m} = \frac{1}{4}$$

Using Eqs. 19 and 20 and the curves of Fig. 10 one can, for each β find the threshold CNR. This is plotted on Fig. 16 as a function of β .

Conclusion

The new method described above for estimating the expected number of clicks per second in the phase locked loops has shown itself capable of giving results that agree reasonably well with those obtained experimentally for the first order loop without modulation* . Methods to extend the method to cases with modulation are being developed, and the method can be easily applied to loops of any order.

The great advantage of the method over a direct digital computer estimation of the expected number of clicks per second using gaussian noise, is computer time. The direct method would use hours if not days of computer time to get a reasonable estimate at one value of carriers to noise ratio. In addition this method provides a closed expression, in terms of carrier to noise ratio of the expected number of clicks per second. It can also be used to compare various PLL designs.

The method is expected to be applicable to other problems too, such as FM with feedback.

* The only experimental results immediately obtainable were taken with a critically coupled coil rather than a cascade of two single tuned coils which was considered in this report. The discrepancy between experimental and theoretical results shown in Fig. 10 for the gain of 1 , can be attributed to the different IF filter, and also to the straight line approximations in Fig. 4, 5, and 6.

$$V_i(t) = \cos \omega_0 t - y(t) \cos \omega_0 t + x(t) \sin \omega_0 t$$

$$= \sqrt{(1-y)^2 + x^2} \cos \left[\omega_0 t + \tan^{-1} \left(\frac{x}{1-y} \right) \right]$$

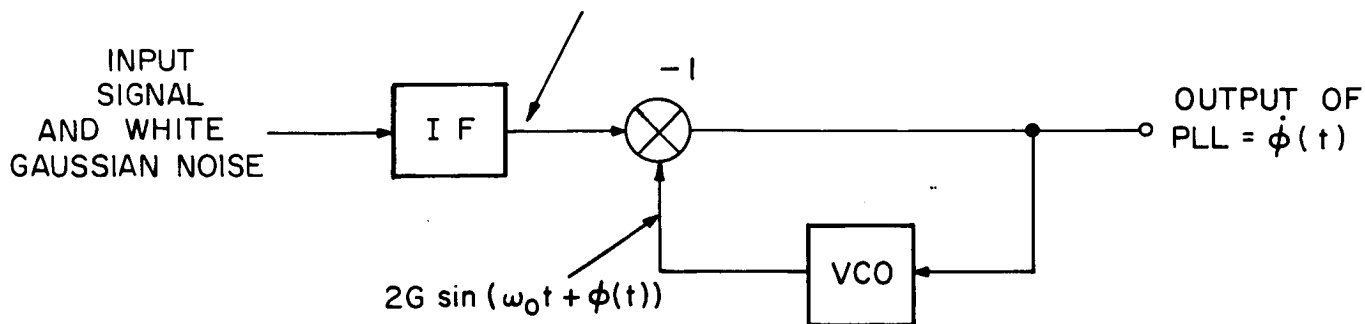


Fig. 1 The PLL

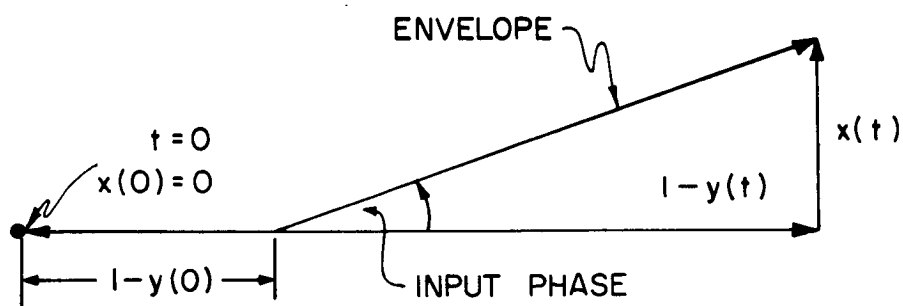


Fig. 2 Phase Design of Input Signal and Noise

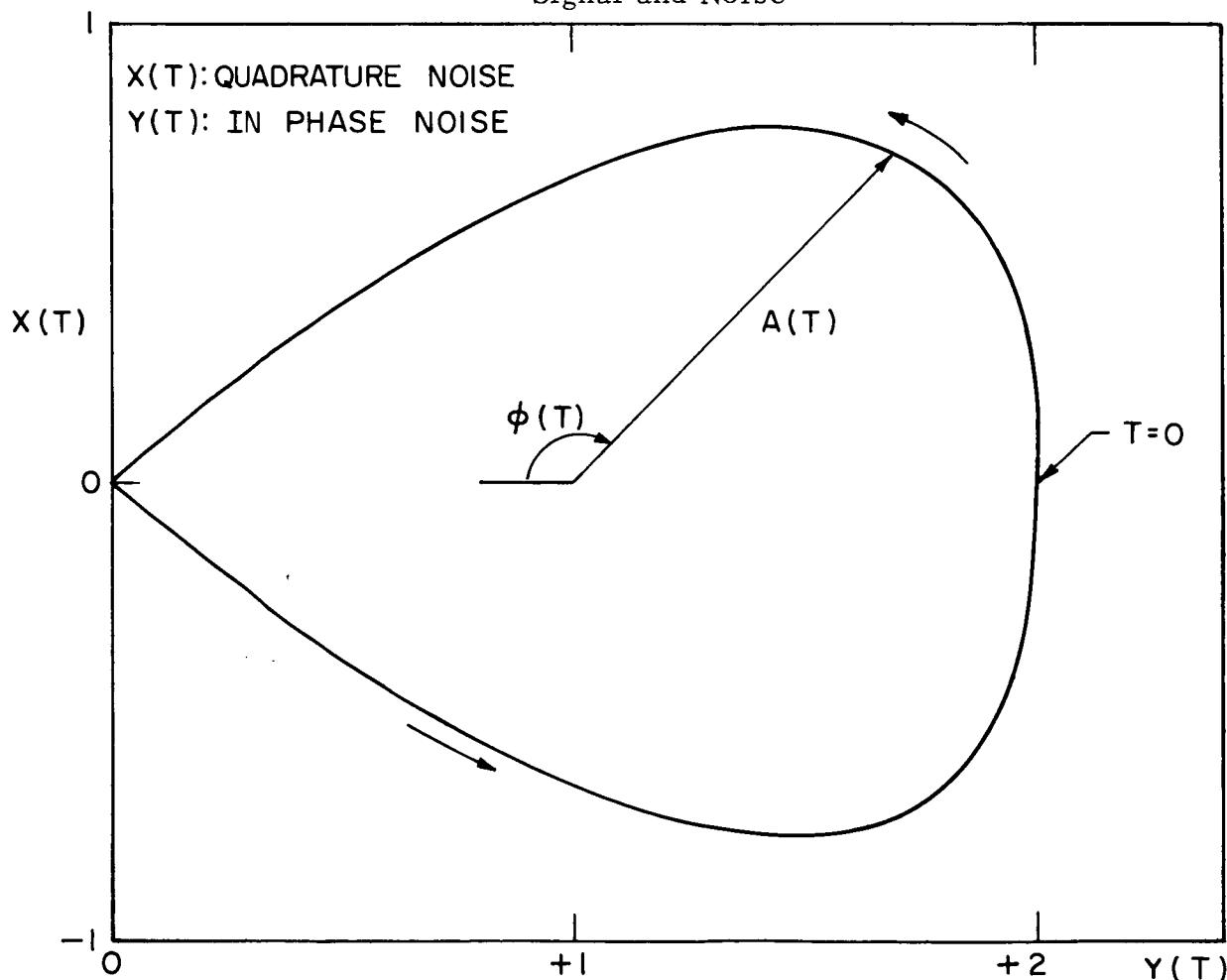


Fig. 3 Polar plot of most probable noise

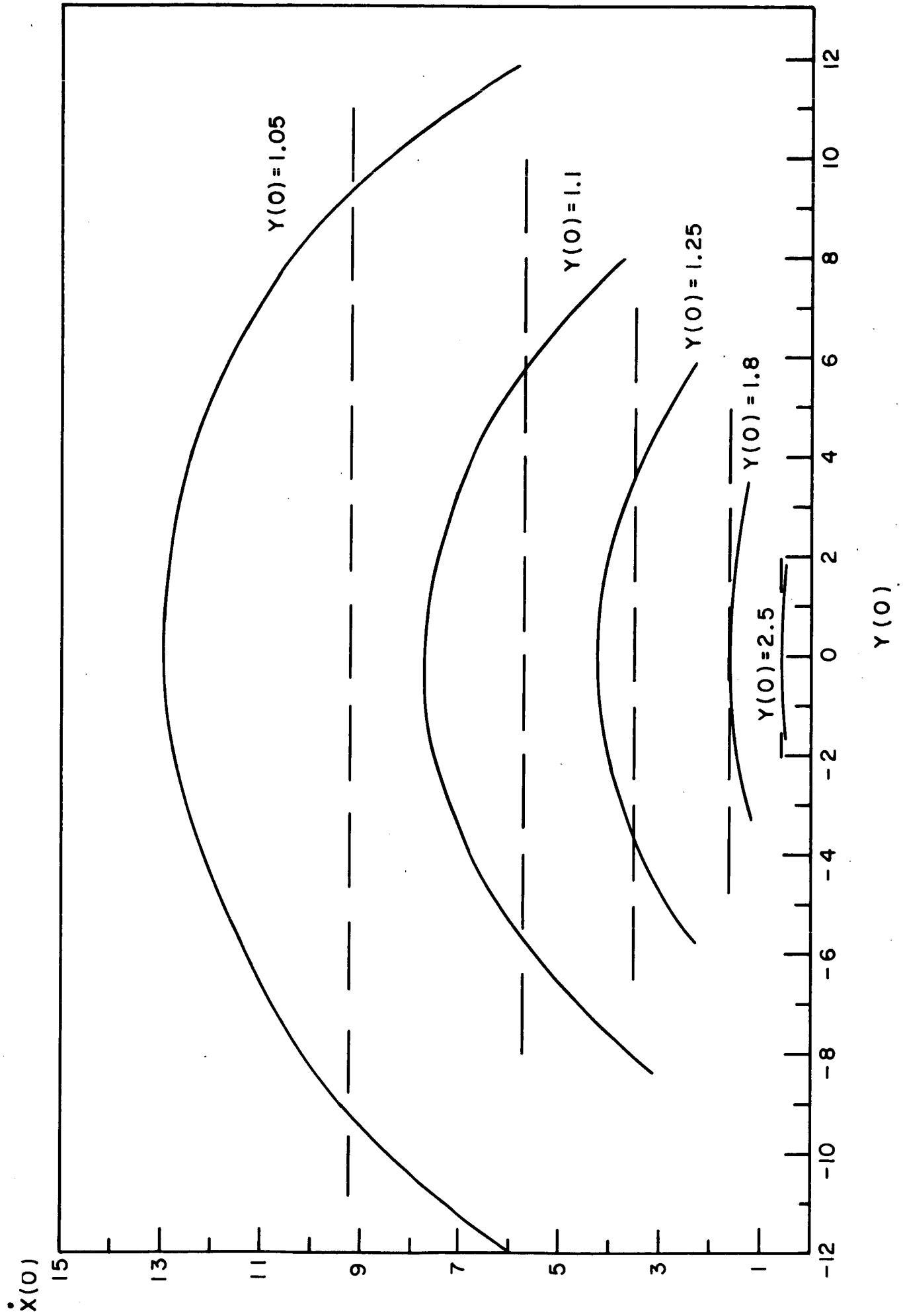


Fig. 4 Click Boundary Surface Loop Gain $G : G = 1$.

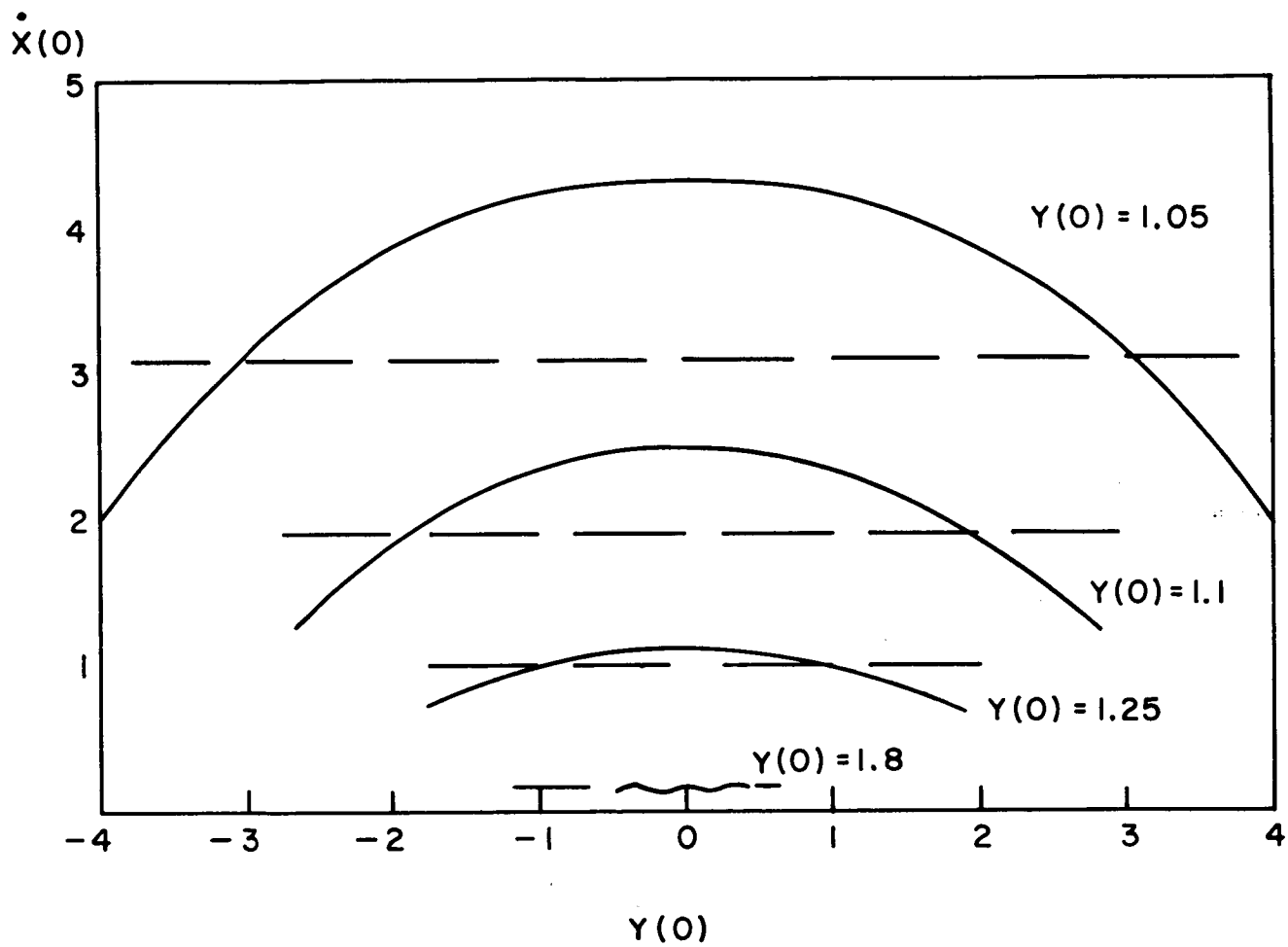


Fig. 5 Click Boundary Surface Loop Gain G : $G = 3$.

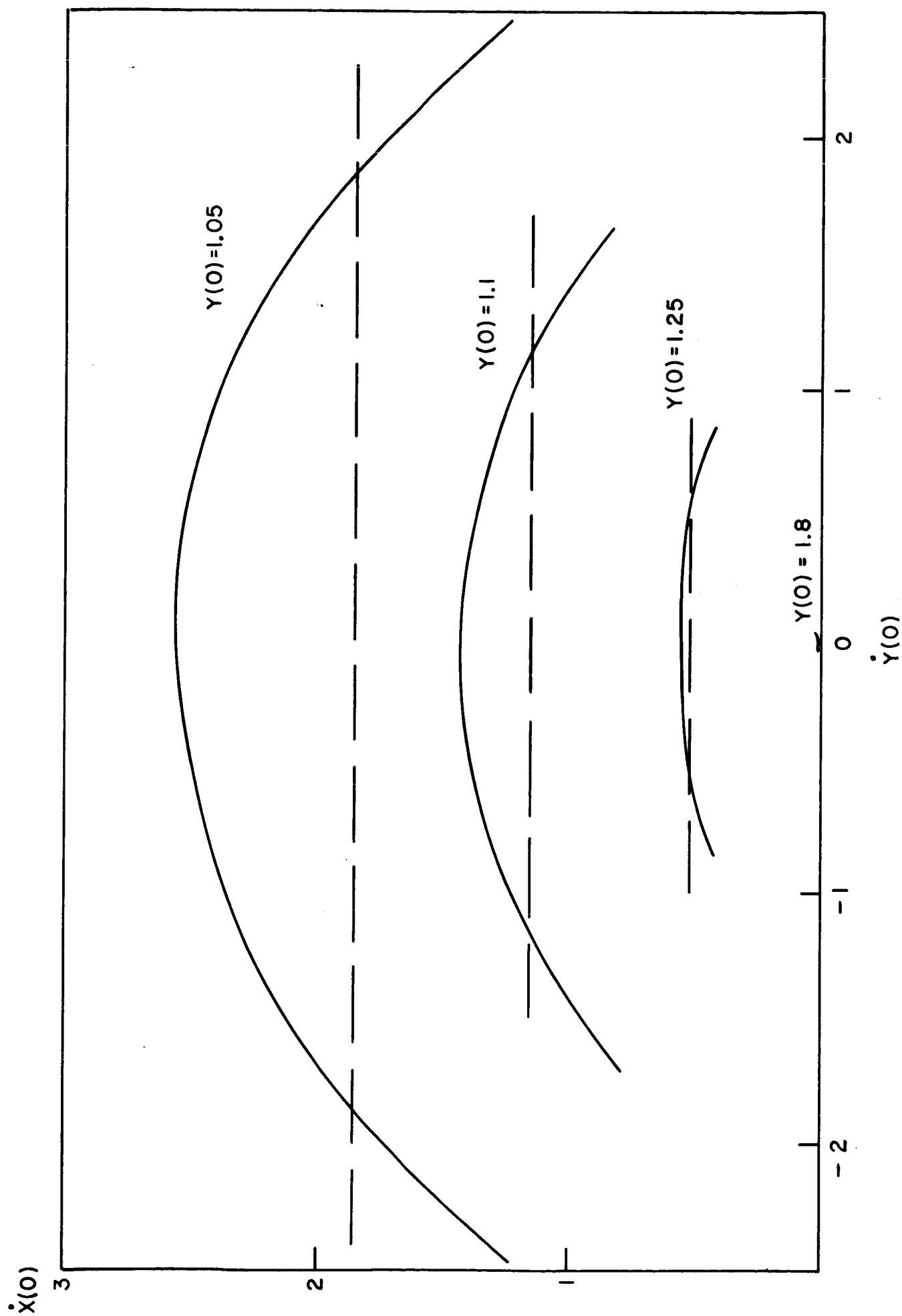


Fig. 6 Click Boundary Surface Loop Gain G : $G = 5$.

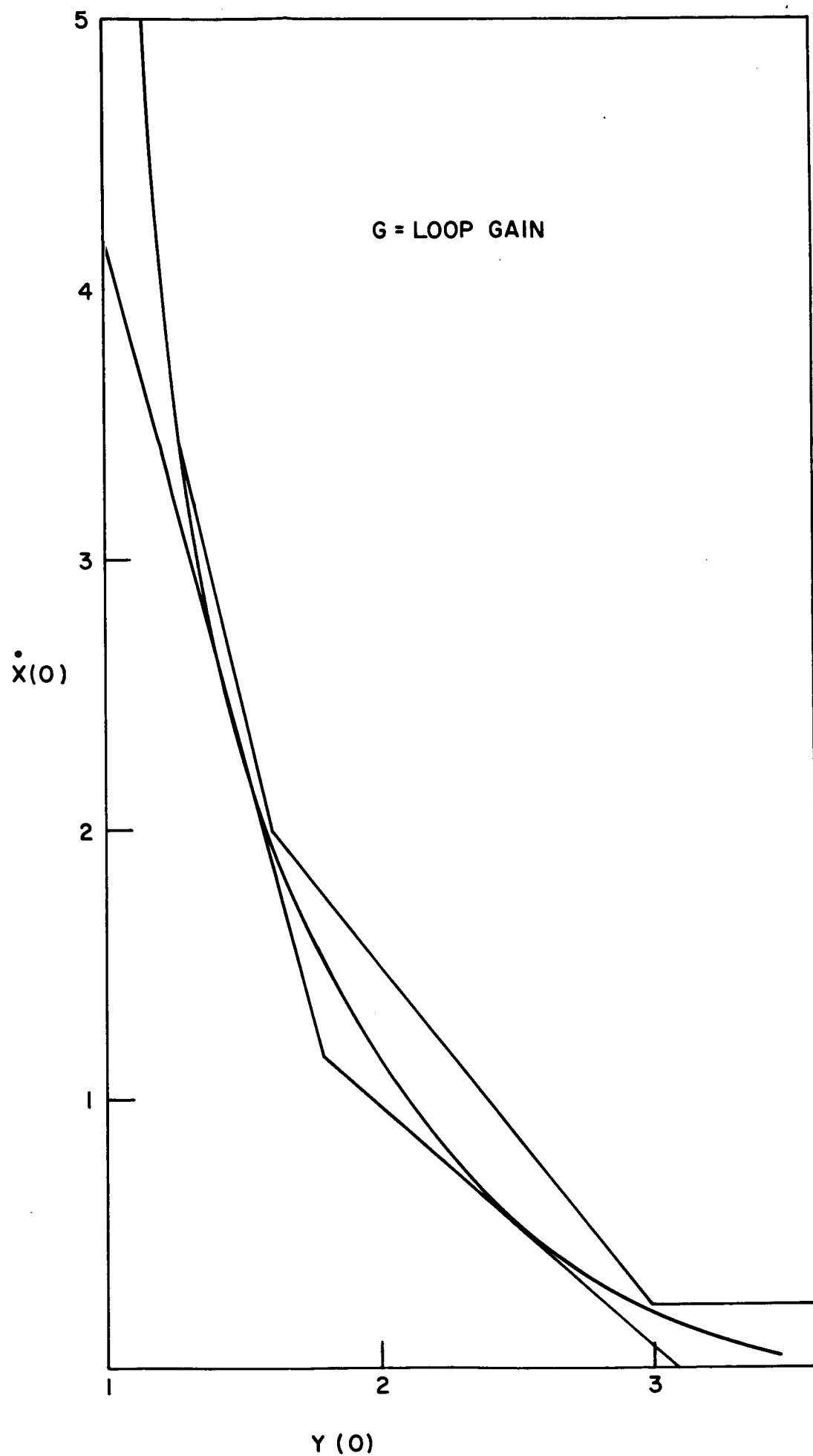


Fig. 7 Approximation to Click Boundary $G = 1$.

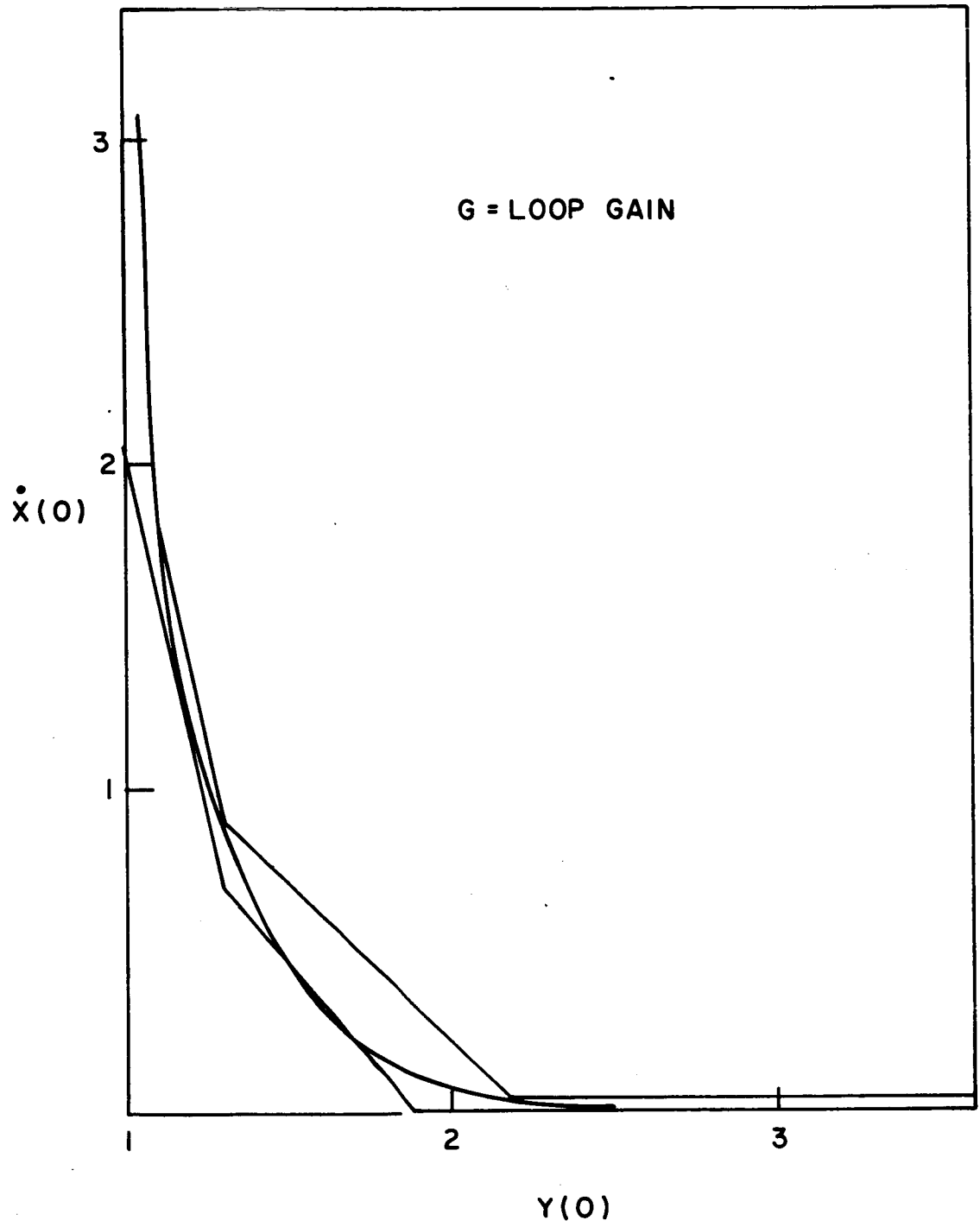


Fig. 8 Approximation to Click Boundary $G = 3$.

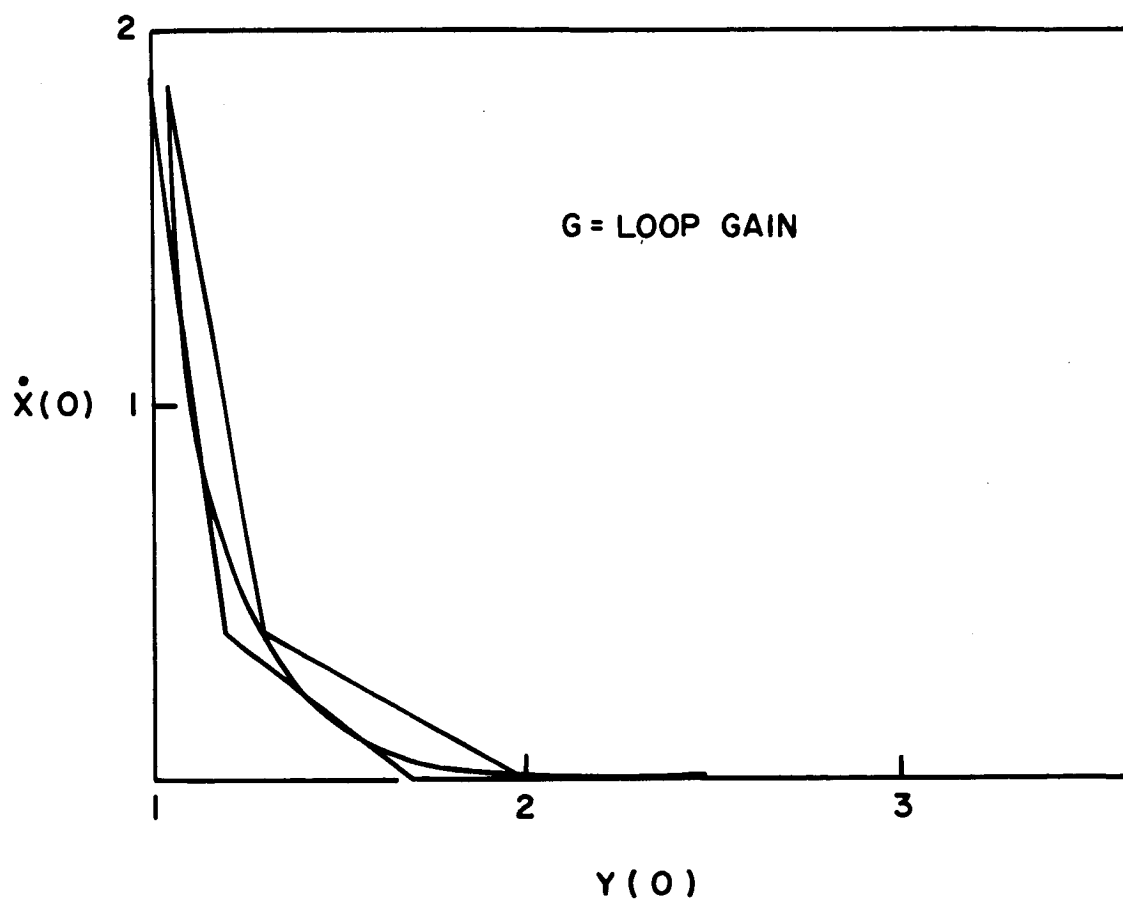


Fig. 9 Approximation to Click Boundary $G = 5$.

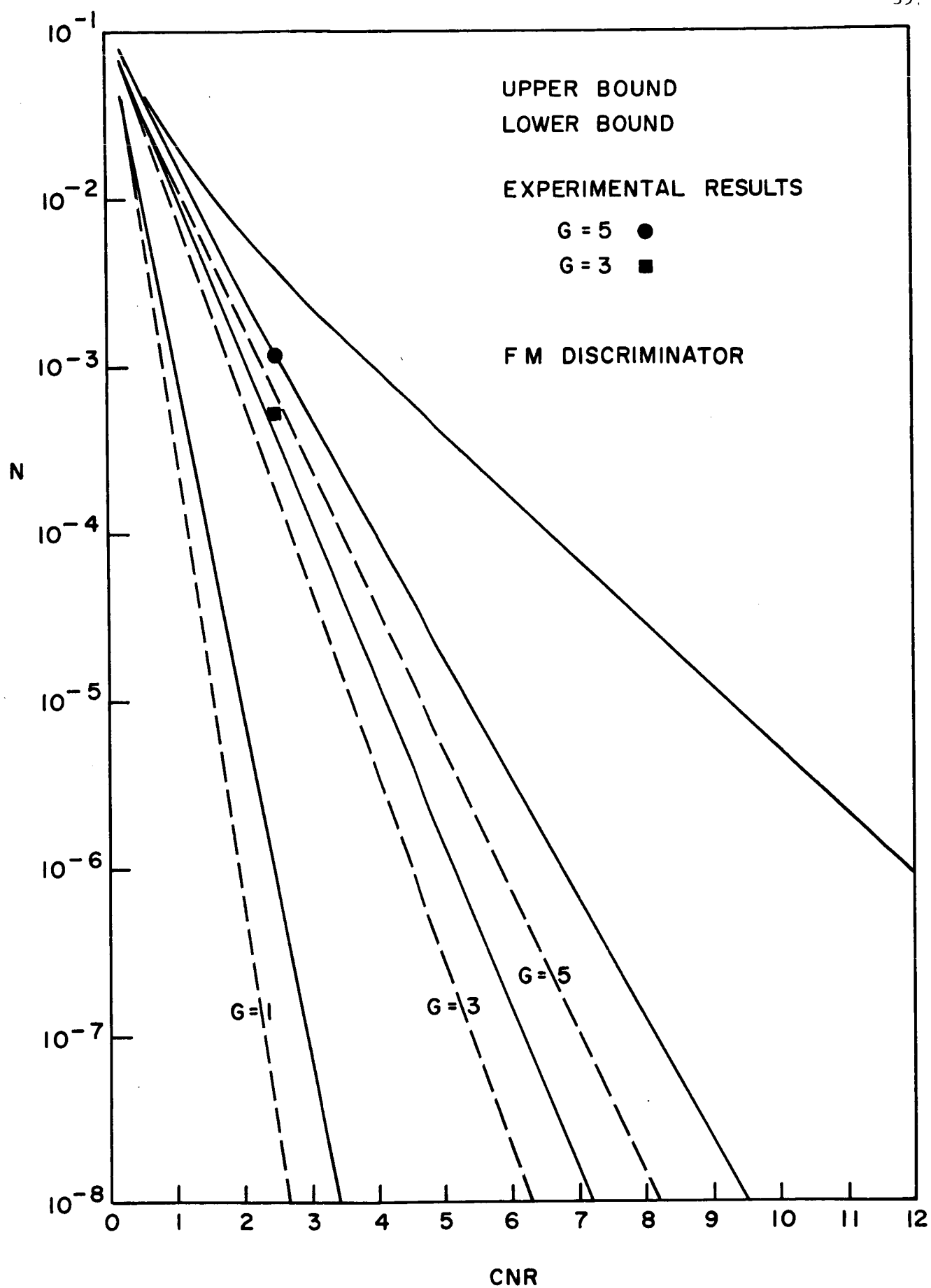


Fig. 10 Average Number of Clicks per Second Vs. Carrier-to-Noise Ratio

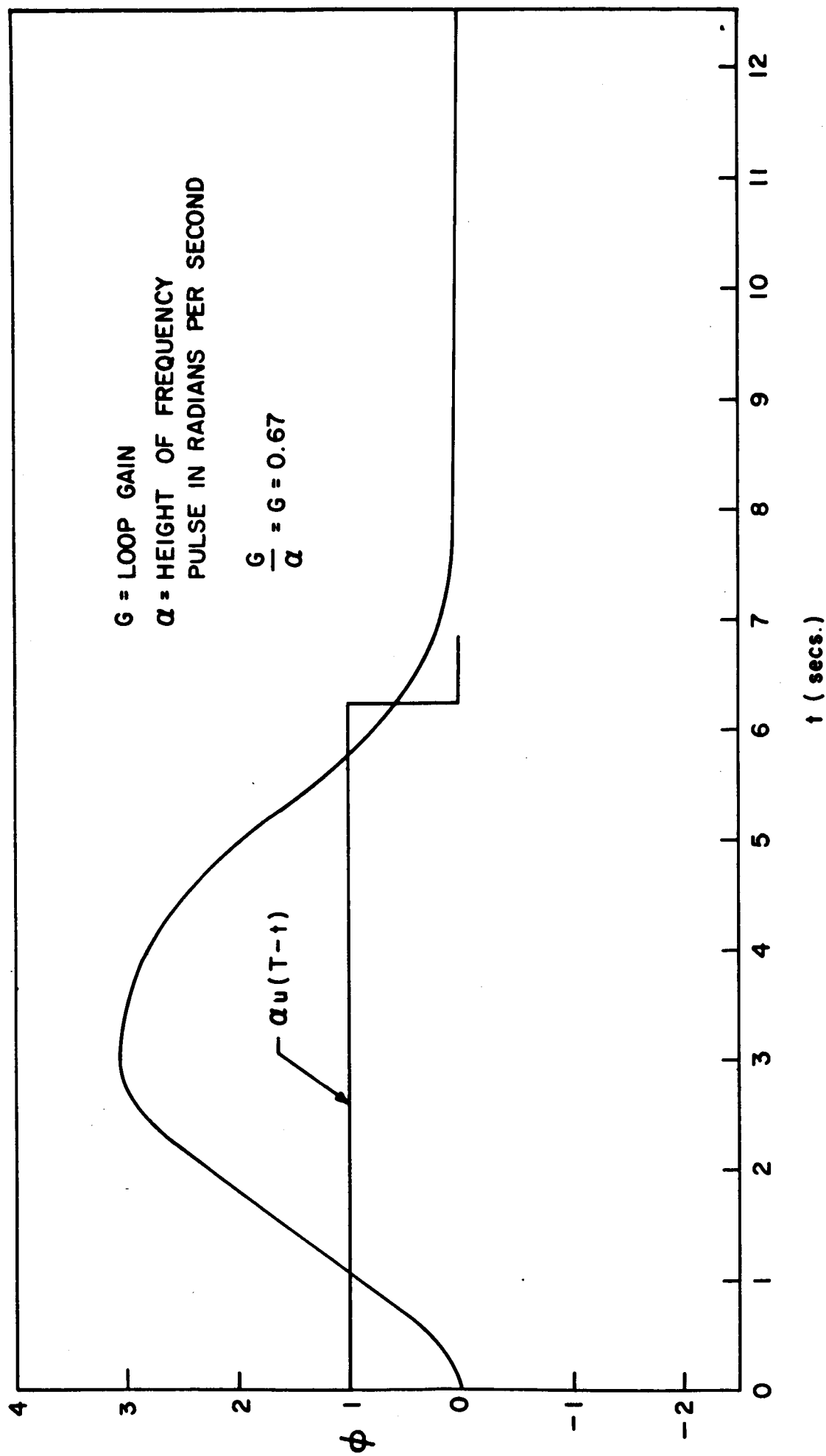


Fig. 11 Output of First Order PLL with Frequency Input

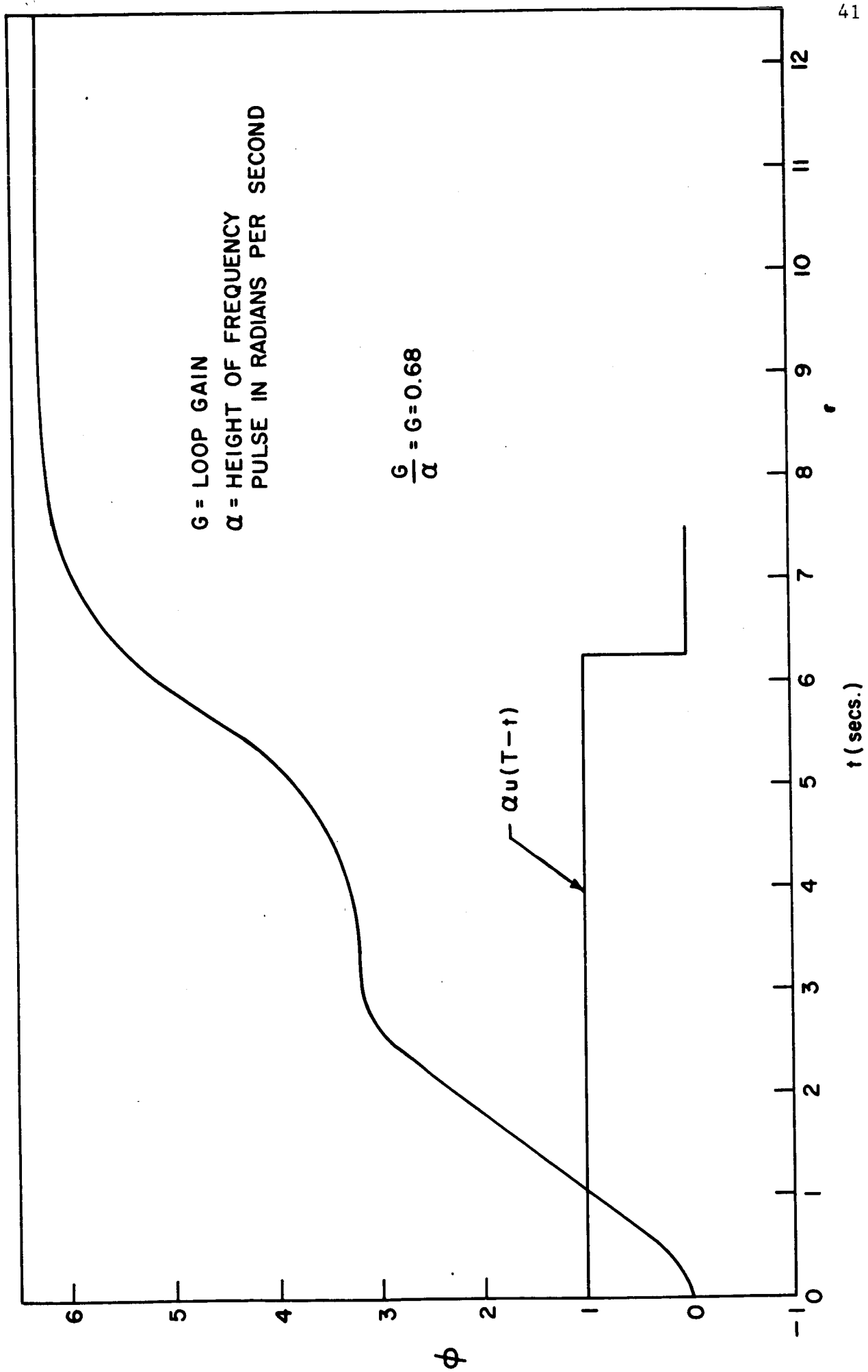


Fig. 12 Output of First Order PLL with Frequency Pulse Input

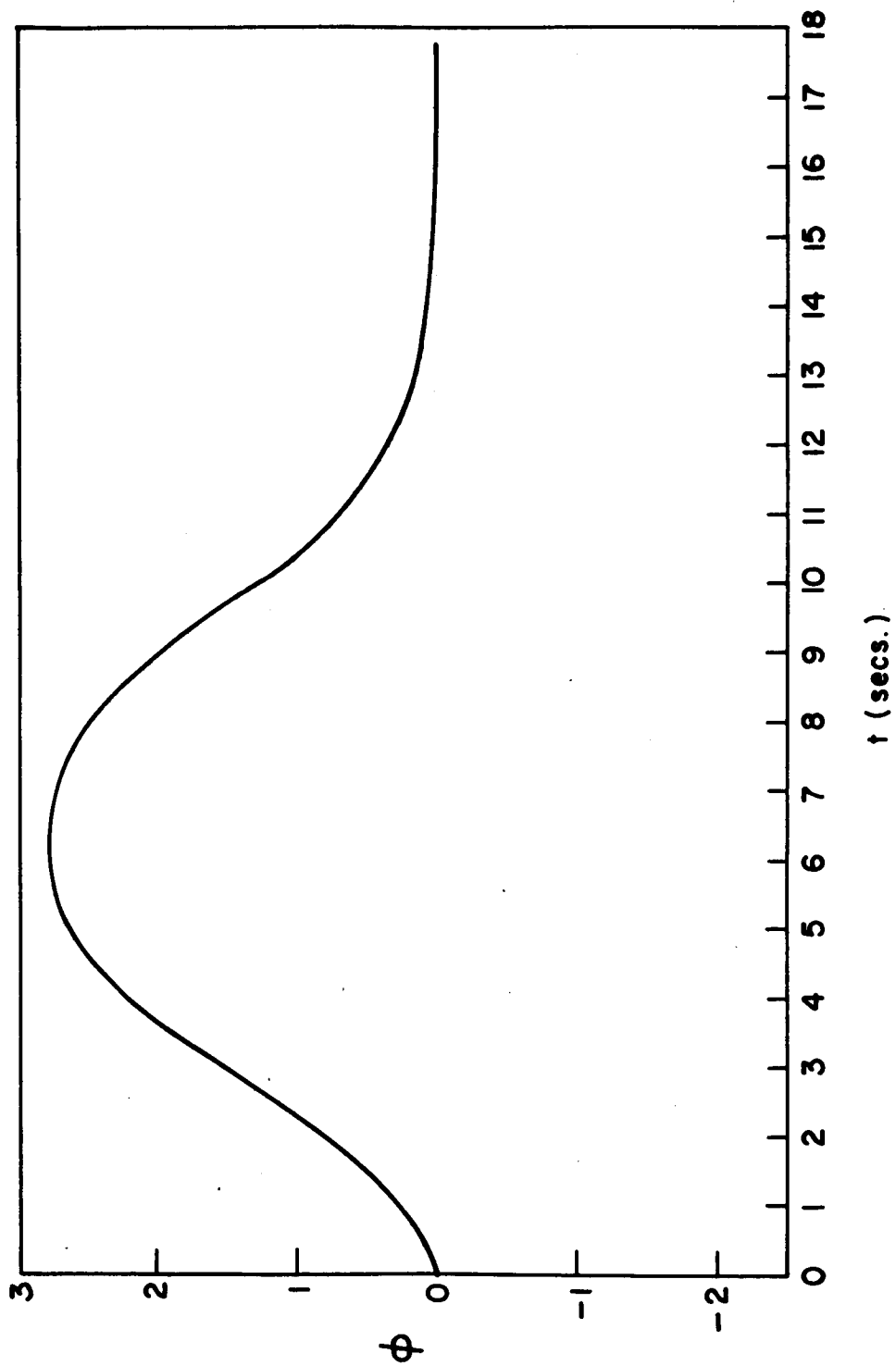
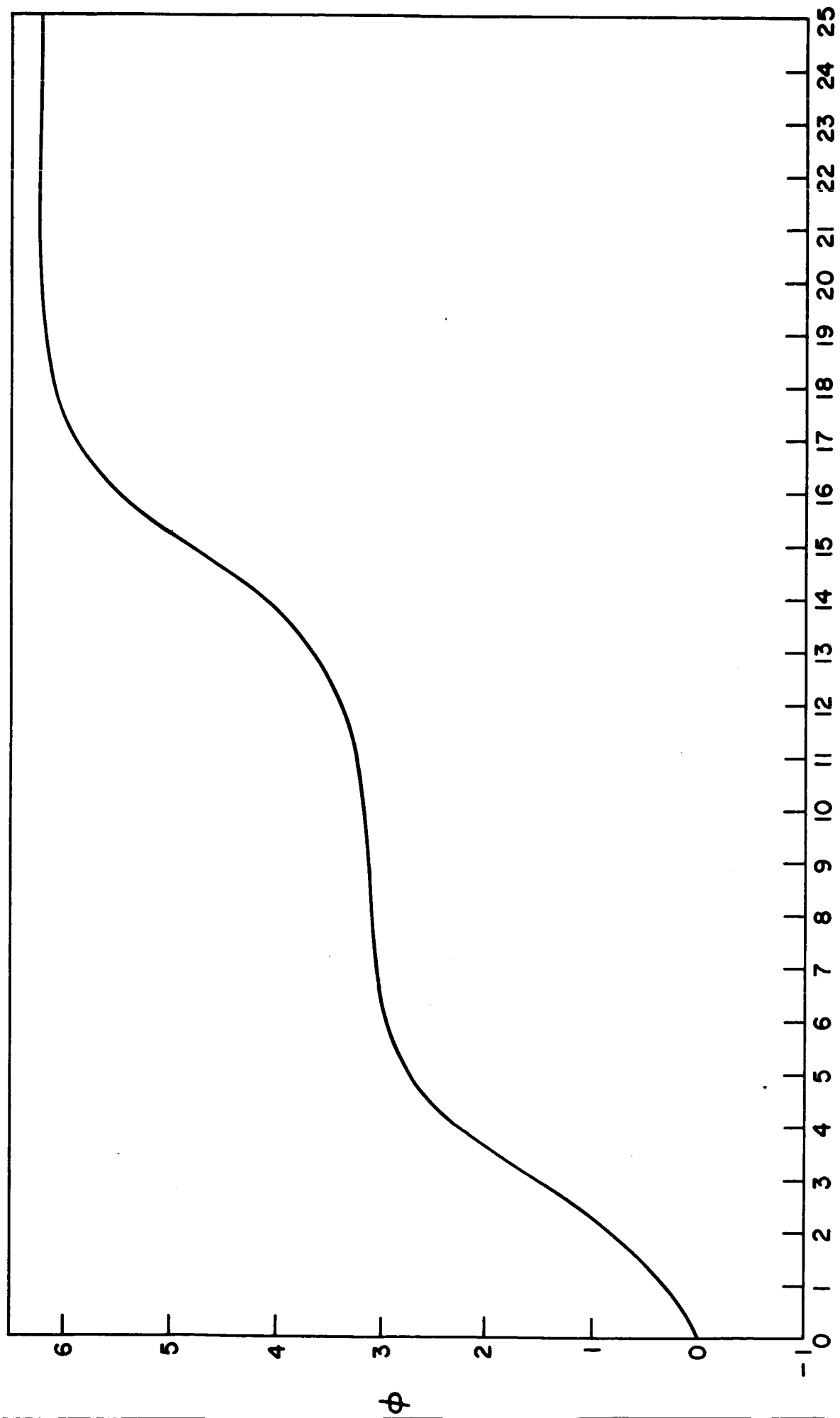


Fig. 13 Output of PLL with Most Probable Noise and Gain of 0.725.



t (secs.)

Fig. 14 Output of PLL with Most Probable Noise and Gain of 0.75

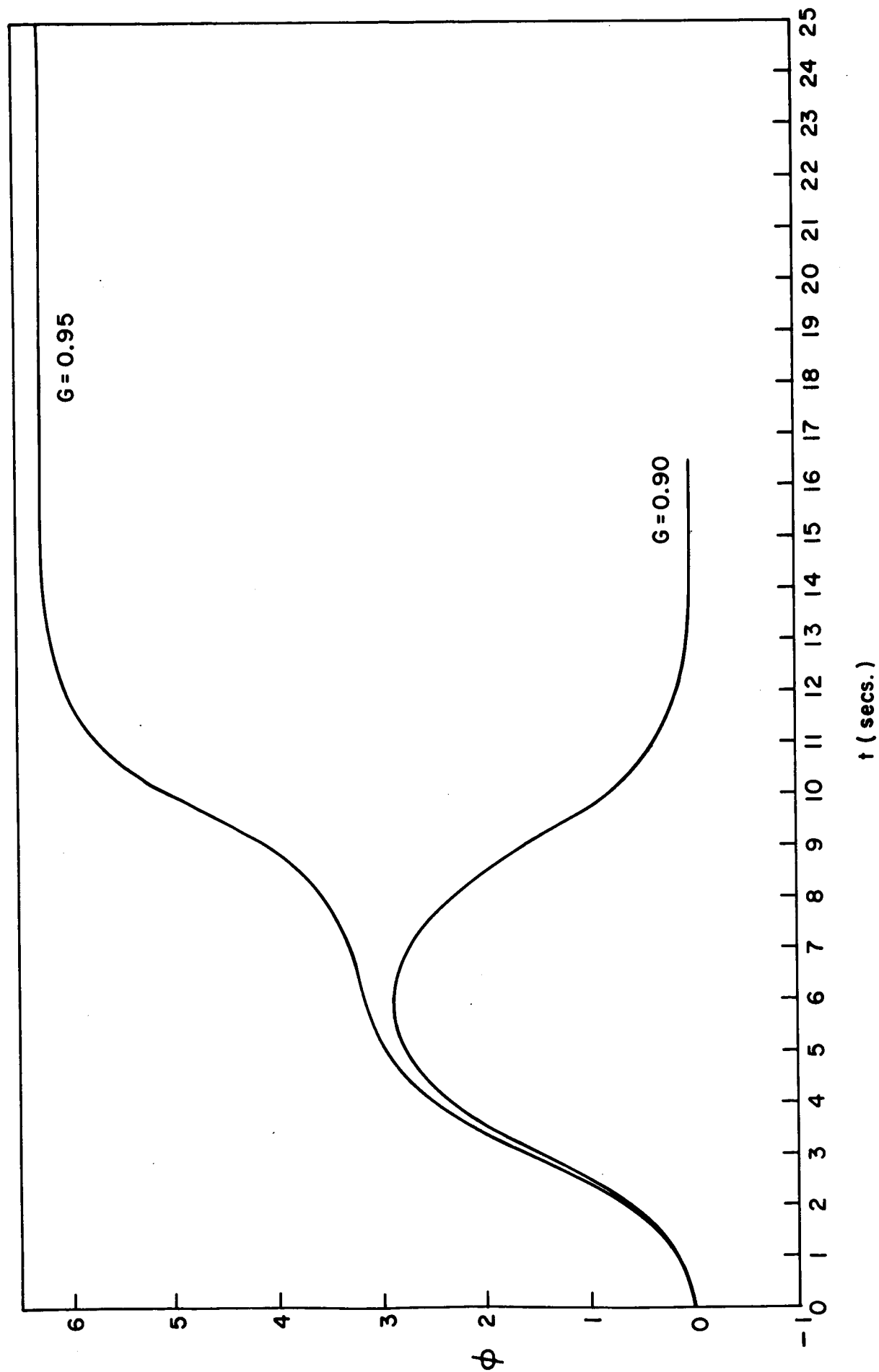


Fig. 15 Output of PLL with Perturbed Most Probable Noise and a Gain of 0.90 and 0.95.

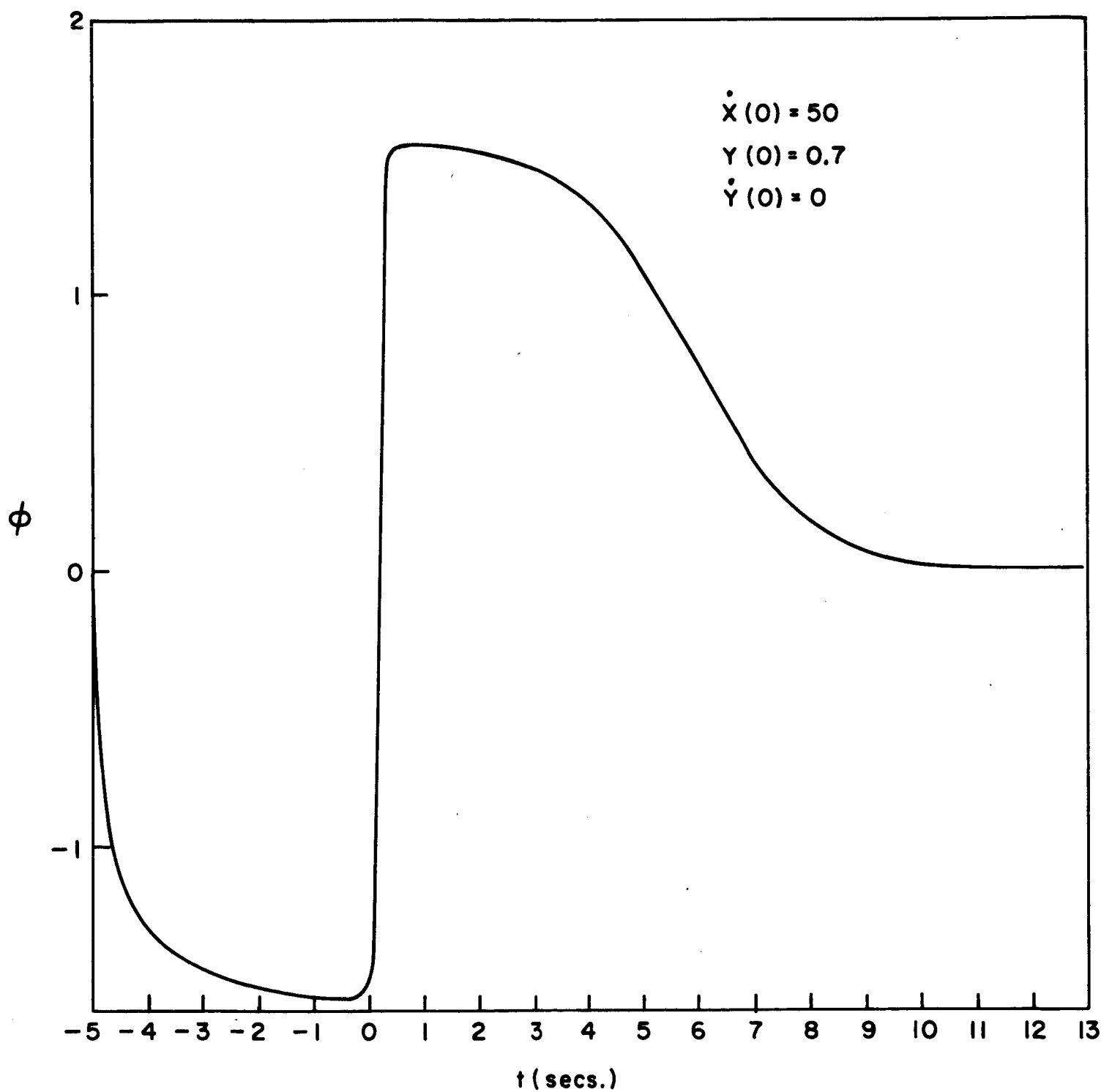


Fig. 16 Output of PLL with Most Probable Noise
(Doublet) $G = 3$

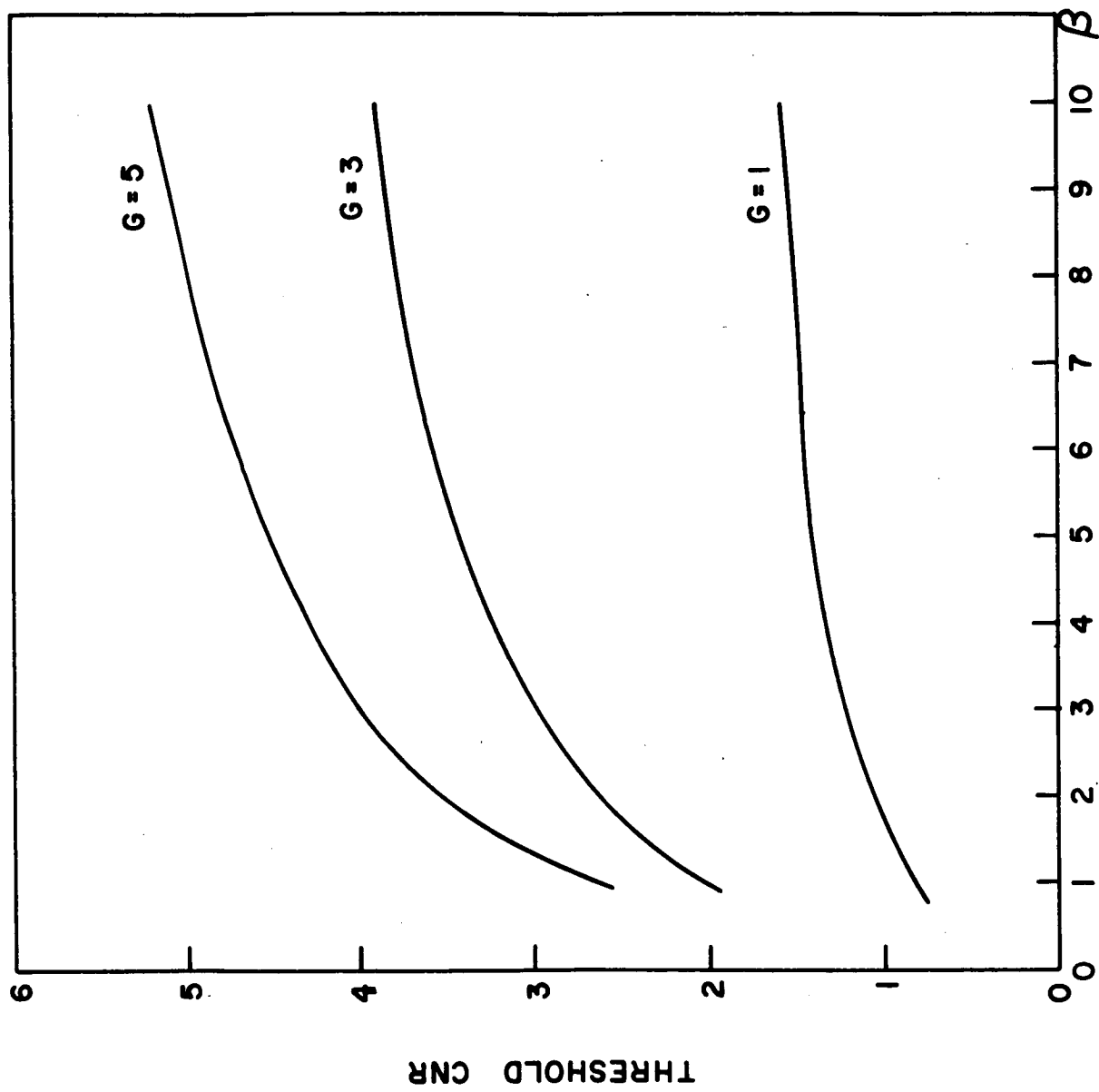


Fig. 17 Threshold CNR vs β

I-B.1b.

Pulse Model for Input NoiseIntroduction

In this section a new model for a carrier plus narrow band-noise is presented and with its aid the expected number of "clicks" per second occurring at the output of a first order Phase-Locked Loop (PLL) excited by a narrow-band noise coupled carrier is determined. The expected number of "clicks" (or "cycle slipping events" as they are called when the PLL is used for tracking purposes) per second is determined for the PLL alone and also for the PLL preceded by a limiter. For both cases corresponding experimental data are presented which indicate the extreme accuracy of results obtained with the use of the new model.

The motivation for the model is based on representing the input signal e_o to the PLL in the form

$$e_o(t) = A \cos \omega_o t + n(t) = a(t) \cos [\omega_o t + \psi(t)] \quad (1)$$

where ω_o is the carrier frequency, A is the carrier amplitude, and $n(t)$ is the noise appearing at the output of a symmetric narrow-band filter centered about ω_o . The function $\psi(t)$ represents the instantaneous phase of the PLL input which the loop tries to track. It is well known (1) however, that $\psi(t)$ contains steps of $\pm 2\pi$ which result in output "clicks" when $e_o(t)$ is processed by a limiter-discriminator. Consequently, if the PLL tracks $\psi(t)$ exactly it would have the same frequency "clicks" as the discriminator, and the problem would be solved. The loop, however, does not track all of the steps of $\pm 2\pi$ in $\psi(t)$ and thus has fewer output "clicks" than the limiter-discriminator and consequently an improved threshold performance. The problem which must be solved is therefore the determination of how many steps of $\pm 2\pi$ the PLL tracks. The solution to this problem yields the expected number of cycle slipping events or "clicks" per second in the PLL.

It should be noted that the point of view presented here is somewhat unconventional. Usually, a cycle slipping event is thought to occur when the PLL loses lock in the presence of noise, In this analysis it is assumed that output "clicks" occur when the loop tracks the noise-induced steps of $\pm 2\pi$ in $\psi(t)$ and that no output "click" occurs when the loop slips or "loses lock" during such a step in $\psi(t)$. Experimentally, the writer has found this point of view to be justified. By placing the same carrier plus narrow band noise into both a first order PLL and a limiter-discriminator and observing the demodulated outputs of both devices simultaneously, he observed that all noise "clicks" appearing at the output of the PLL also occurred at the discriminator output, whereas many of the discriminator "clicks" did not appear at the output of the PLL. These observations were made for a significantly large number of PLL closed loop bandwidths.

It should not be concluded that the PLL never contains an output "click" that would not appear in the output of a limiter-discriminator. If a modulated carrier plus noise is applied to the input of the PLL and the frequency deviation of the carrier is increased, or the carrier is detuned then a large number of signal induced "clicks" appear at the output of the PLL. This makes sense since the modulated carrier moves the operating point of the PLL close to the edge of the "hold in range" where it becomes easy for additive noise to cause a "loss of lock" with its resultant output "click". It is convenient to refer to the signal induced "clicks" as "clicks" of the second kind. The "clicks" which result from the PLL tracking steps of $\pm 2\pi$ appearing at the input are referred to a "clicks" of the first kind.

Although the ultimate problem is to calculate, in terms of the parameters of the PLL, the expected number of both "clicks" of the first and second kind, this paper deals only with a means of calculating the expected number of "clicks" of the first kind.

In the sections to follow the model for the input carrier plus noise is developed and applied to the first order PLL both with and without a limiter to determine the expected number of cycle slipping events or output "clicks" per second. It is found that these "clicks" are a function of the ratio of the closed loop bandwidth of the PLL to the radius of gyration $\gamma^{(2)}$ of the narrowband input noise. Several curves of expected number of output "clicks" vs. the ratio of closed loop bandwidth to input noise radius of gyration are plotted with the input carrier to noise ratio as a parameter. In particular it is shown that when the closed loop bandwidth remains fixed, the output of the PLL and the limiter-discriminator are identical. Finally, experimental data are presented which verify that the model made to characterize the input carrier plus noise is indeed a valid one. In all cases agreement within a few percent between theory and experimentation is observed.

II. Model for Input Carrier Plus Noise

A typical plot of $a(t)$ and $\dot{\psi}(t)$ of equation (1) vs. time takes the form shown in figure 1. At the times where $\psi(t)$ contains steps of $\pm 2\pi$, $\dot{\psi}(t)$ consists of pulses of area $\pm 2\pi$. To model these signals the following assumptions are made.

1) A step of $\pm 2\pi$ in $\psi(t)$ from $2n\pi$ to $2(n\pm 1)\pi$ occurs in such a fashion that $\dot{\psi}(t)$ may be closely approximated by a truncated Gaussian time function of area $\pm 2\pi$ given by

$$\dot{\psi}(t) = \pm \dot{\psi}(t_i) e^{-\frac{(t-t_i)^2 [\dot{\psi}(t_i)]^2}{4\pi}}$$

$$= 0$$

$$|t-t_i| \leq \frac{4\sqrt{2\pi}}{\dot{\psi}(t_i)}$$

$$|t-t_i| > \frac{4\sqrt{2\pi}}{\dot{\psi}(t_i)}$$

where t_1 is the time at which $\psi(t_1) = 2(n \pm \frac{1}{2})\pi$. Because of the truncation the area of $\dot{\psi}(t)$ is slightly less than $\pm 2\pi$; however this difference has no effect on any subsequent calculations.

2) The height of all input pulses are identical and equal to twice the radius of gyration γ of the input noise, i. e., $\dot{\psi}(t_1) = 2\gamma$. This choice of pulse height is not based upon a rigorous theoretical analysis but is rather the result of a lengthy experimental study during which it was found that such a pulse description gave accurate results in all situations encountered. Several other models describing $\dot{\psi}(t)$ were initially employed which took into account differences in input "click" height. These, however, not only required the use of a computer, but also yielded less accurate results.

3) Prior to the occurrence of a step of $\pm 2\pi$ in $\psi(t)$ the loop is tracking $\psi(t)$ with zero error signal.

4) $a(t)$ remains essentially constant at $a(t_1)$ during the occurrence of the pulse in $\dot{\psi}(t)$. This assumption makes sense since variations in $a(t)$ are found (both theoretically and experimentally) to be slow compared with the duration of a pulse in $\dot{\psi}(t)$.

Based upon the above assumptions, the model for $a(t)$ and $\dot{\psi}(t)$ during a step of 2π in $\psi(t)$ is shown in figure 2. With this model one may readily determine whether or not an input step of $\pm 2\pi$ in phase is tracked by the PLL. If it is tracked an input "click" is transmitted to the output. Only if a number of such steps in phase are not tracked can the PLL show any threshold improvement. In the following section the response of the first order PLL to the model of figure 2 is obtained and the conditions under which a cycle slipping event (a PLL output "click") occurs are determined.

III Response of First Order PLL

To determine the output of the PLL, a loop of the form shown in figure 3 is assumed. The second harmonic rejection filter removes terms in the vicinity of $2\omega_o$ such that the equation for the loop takes the form

$$\begin{aligned}\dot{\varphi}(t) &= \frac{a(t)B}{2} \sin [\psi(t) - \varphi(t)] \\ &= \frac{a(t)}{A} \omega_L \sin [\psi(t) - \varphi(t)]\end{aligned}\quad (2)$$

where A is the carrier amplitude and $\omega_L = BA/2$ is the closed loop bandwidth or equivalently the "hold in range" of the loop when no noise appears at the input. By defining $\theta(t) = \psi(t) - \varphi(t)$ and substituting into (2) one obtains

$$\dot{\theta}(t) + \frac{\omega_L a(t)}{A} \sin \theta(t) = \dot{\psi}(t) \quad (3)$$

During the occurrence of a $\pm 2\pi$ step in $\psi(t)$, use of the model of figure (2) yields

$$\dot{\theta}(t) + \frac{\omega_L a(t_i)}{A} \sin \theta(t) = \pm \dot{\psi}(t_i) \epsilon \frac{-[t-t_i]^2 [\dot{\psi}(t_i)]^2}{4\pi}$$

By defining $X = \frac{\dot{\psi}(t_i) [t-t_i]}{\sqrt{2\pi}}$ and $\alpha = \frac{\sqrt{2\pi} \omega_L a(t_i)}{A \dot{\psi}(t_i)}$ equation (4) takes

the normalized form given by

$$\pm \sqrt{2\pi} \epsilon \frac{-X^2}{2} = \alpha \sin \theta'(X) + \frac{d\theta'(X)}{dX} \quad (5)$$

where $\theta'(X) = \theta \left(\frac{\sqrt{2\pi} X}{\dot{\psi}(t_i)} + t_i \right) = \theta(t)$.

For small values of α it is apparent that

$$\theta'(X) \approx \pm (2\pi) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^X e^{-\frac{\tau^2}{2}} d\tau \right) \quad (6)$$

which increases by 2π during the step of 2π in $\psi(t)$; hence $\varphi(t) = \psi(t) - \theta(t)$ does not experience a step of $\pm 2\pi$ and $\dot{\varphi}(t)$ (the output of the PLL) does not contain a "click". For large values of α on the other hand, $\theta'(X)$ is given by

$$\theta'(X) \approx \pm \sin^{-1} \left[\frac{\sqrt{2\pi} e^{-\frac{X^2}{2}}}{\alpha} \right] \approx \pm \frac{\sqrt{2\pi}}{\alpha} e^{-\frac{X^2}{2}} \quad (7)$$

It is clear that $\theta'(X)$ given by equation (7) experiences no step of 2π ; hence $\varphi(t)$ in this case does contain a step of 2π yielding an output "click" in the PLL.

An exact computer solution of equation (5) indicates that for all values of α greater than 1.31, the PLL does not "lose lock" and a "click" appears in the output.

The values of $\alpha \geq 1.31$ correspond to

$$\frac{\omega_L a(t_i)}{A \dot{\psi}(t_i)} = \frac{\omega_L a(t_i)}{2 A \gamma} \geq 0.52 \quad (8)$$

where γ is the radius of gyration of the input narrow-band noise. It is apparent that output "clicks" occur only when $a(t_i)$ is large during the occurrence of a noise induced step of $\pm 2\pi$ in $\psi(t)$. In general, however, since $a(t)$ during a step of $\pm 2\pi$ in $\psi(t)$ is usually small, the PLL tends to slip by most of the "clicks" in $\psi(t)$ which would appear at the output of a limiter-discriminator. For this reason the PLL exhibits threshold extension properties.

It is apparent that the number of output clicks is further decreased as ω_L decreases relative to the radius of gyration of the input noise which is closely related to the bandwidth of the narrow-band input filter.

In the following section, using Rice's model for a "click" combined with the result from the solution of equation 5, an expression for the expected number of "clicks" per second at the output of the PLL is determined.

IV Expected Number of PLL "Clicks"

For a "click" to occur at the output of the PLL two conditions must be satisfied:

- 1) a step of $\pm 2\pi$ must occur in $\psi(t)$ and
- 2) $\frac{\omega_L a(t_i)}{2A\gamma} \geq 0.52$ where t_i is the time $\psi(t)$ crosses $(2n \pm 1)\pi$ and γ

is the radius of gyration of the input narrow band noise. The probability of both conditions occurring in a small time interval Δt which contains t_i yields the probability $p_o \Delta t$ of a PLL output "click" occurring in Δt , whereas p_o is equal to the expected number of PLL output "clicks" per second if one follows Rice's argument⁽⁴⁾ that the times t_i are Poisson distributed.

The two conditions necessary for a PLL "click" may be expressed in a more conventional form by writing the input noise to the PLL in the form

$$n(t) = \rho(t) \cos [\omega_o t + \theta(t)],$$

and the input carrier plus noise in the form

$$\begin{aligned} e_o(t) &= A \cos \omega_o t + \rho(t) \cos [\omega_o t + \theta(t)] \\ &= \operatorname{Re} e^{j\omega_o t} [A + \rho(t) e^{j\theta(t)}] \\ &= \operatorname{Re} e^{j\omega_o t} [a(t) e^{j\psi(t)}] \end{aligned} \tag{9}$$

Figure 4 shows the phasor diagram for the terms in brackets in equation (9).

If one follows Rice's additional argument⁽⁵⁾ that a step of $\pm 2\pi$ in $\psi(t)$ occurs in the interval Δt (containing t_i) if, during Δt

$$\begin{aligned} & (\pi \leq \theta(t) < \pi + \Delta\pi, \dot{\theta}(t) > 0 \text{ (positive step), and } \rho(t) > A) \\ & (\pi \leq \theta(t) < \pi + \Delta\pi, \dot{\theta}(t) \leq 0 \text{ (negative step), and } \rho(t) > A) \end{aligned} \quad (10)$$

and, in addition, notes that (8) is equivalent to

$$\rho(t_i) \geq A \left[1 + \frac{(1.04)\gamma}{\omega_L} \right] \quad (11)$$

(when $\theta(t) = \theta(t_i) = \pm\pi$, $a(t_i) = \rho(t_i) - A$), one may write the probability of a PLL output "click" occurring in Δt in the form

$$p_o \Delta t = \Delta t \int_{-\infty}^{\infty} \left[1 + \frac{1.04\gamma}{\omega_L} \right] \int_{-\infty}^{\infty} |\dot{\theta}(t_i)| p \left\{ \rho(t_i), \theta(t_i) = \pi, \dot{\theta}(t_i) \right\} d\dot{\theta} d\rho \quad (12)$$

where $p[\rho, \theta, \dot{\theta}]$ is the probability density function of ρ, θ , and $\dot{\theta}$. Consequently, if the t_i 's are Poisson distributed, the expected number of PLL output "clicks" per second N^+ is given by $N^+ = p_o$.

In addition, if the narrow band input noise is assumed to be stationary gaussian noise with a power spectrum symmetric about ω_o , $p(\rho, \theta, \dot{\theta})$ takes the form

$$p(\rho, \theta, \dot{\theta}) = \frac{\rho^2}{\gamma^2 (2\pi N)^{3/2}} e^{-\frac{\rho^2}{2N}} e^{-\frac{\rho^2 \dot{\theta}^2}{2\gamma^2 N}} \quad (13)$$

where $N = E[n^2(t)]$ is the total noise input power to the PLL. As a result the expected number of PLL output "clicks" per second is given by

$$N^+ = \left(\frac{\gamma}{\pi}\right) \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \left[\frac{A}{\sqrt{N}} \left(1 + \frac{1.04\gamma}{\omega_L} \right) \right] \quad (14)$$

$$= \frac{\gamma}{\pi} \operatorname{cerf} \left[\frac{A}{\sqrt{N}} \left(1 + \frac{1.04\gamma}{\omega_L} \right) \right]$$

$$\text{where } \operatorname{cerf} Y = \frac{1}{\sqrt{2\pi}} \int_{-Y}^Y e^{-\frac{x^2}{2}} dx$$

With the aid of (14) one may readily calculate N^+ for any given input carrier to noise ratio $\left(\frac{A}{\sqrt{N}}\right)$ once the closed loop bandwidth (ω_L), and the radius of gyration of the input noise are known. It is interesting to note that as ω_L becomes large compared with the radius of gyration of the input noise, N^+ approaches

$$N^+ = \frac{\gamma}{\pi} \operatorname{cerf} \frac{A}{\sqrt{N}} \quad (15)$$

which is exactly the expected number of "clicks" per second obtained at the output of a limiter-discriminator. Numerical substitution into 14, however, indicates that even when $\omega_L = 5\gamma$ a substantially lower number of output "clicks" is obtained from the PLL than from the limiter discriminator.

In the following section a specific numerical example is worked out and a PLL is constructed having the parameters of the example. Experimental results are plotted on the theoretical curves and exceptionally close agreement is obtained, indicating that the model for the input noise has some merit.

V Numerical Example and Experimental Results

An experimental PLL was constructed with a center frequency of 455kHz . A single high frequency transistor switch was used as the loop multiplier, and a Wavetek model 105 variable frequency generator was used as the VCO . This generator was measured to have less than 0.1% total harmonic distortion for frequency deviations of ± 50 kHz about the center frequency. The input noise, obtained from a General Radio Noise Generator (model 1390-B) , was placed through a Collins mechanical filter which has essentially a rectangular pass-band of 13kHz peak to peak bandwidth centered about 455kHz . The input 455kHz carrier was obtained from a General Radio sine wave oscillator model 1310-A .

The complete experimental set up is shown in figure 5. The output of the PLL was placed through a low pass filter and recorded on the storage scope. The low-pass filter removed much of the "grass like" output noise such that the output "clicks" were readily recognizable on the storage scope.

In all measurements the input carrier level was kept constant (well below the limiting dynamic range of any element within the loop) and the input carrier to noise ratio $\frac{A}{\sqrt{2N}}$ was varied by adjusting the output of the noise generator. The closed-loop bandwidth f_L was adjusted by means of the input volume control to the VCO. For each setting of this volume control the "hold in range" and "pull in range" divided by 2 was chosen for f_L .

If one assumes that the noise generator output is flat over the band of frequencies occupied by the mechanical filter, the power spectrum of the input noise to the PLL is rectangular with a radius of gyration given by

$$\gamma = \frac{2\pi (BW)}{2\sqrt{3}}$$

where BW is the bandwidth of the mechanical filter. Consequently, the expected number of PLL output "clicks" per second from (14) is given by

$$N^+ = \frac{BW}{\sqrt{3}} \operatorname{cerf} \frac{A}{\sqrt{N}} \left[1 + \frac{(0.3) BW}{f_L} \right] \quad (16)$$

where $BW = 13 \times 10^3 \text{ Hz}$. N^+ vs. $\left(\frac{BW}{2f_L}\right)$ is plotted in figure 6 with $\frac{A}{\sqrt{2N}}$ as a parameter. Experimentally obtained points are also indicated in figure 6 and show exceptionally close agreement with the theory. Data similar to that presented in figure 6 were also obtained using a double-tuned narrow-band filter preceding the PLL. Agreement between the experimental data and the theoretical curves generated by (14) was again very good.

VI First Order PLL Preceded by a Limiter

It is interesting to determine the expected number of PLL output "clicks" per second when the PLL is preceded by an ideal limiter and the noise model of figure (2) is employed. With a limiter placed after the narrow-band filter the variations in $a(t)$ are removed and the input to the PLL takes the form

$$e_o(t) = A \cos [\omega_o t + \psi(t)] \quad (17)$$

where it is assumed that the limiter constant is adjusted to yield an output of A and higher harmonics of the output are removed by filtering. For this case $a(t_i) = A$ in (8), and it becomes clear that if $\psi(t)$ experiences a step of $\pm 2\pi$ a PLL output "click" occurs if

$$\frac{\omega_L}{\dot{\psi}(t_i)} = \frac{\omega_L}{2\gamma} \geq 0.52 \quad (18)$$

Equation (18) is interesting because it states that for a fixed input filter and a fixed closed loop bandwidth, the expected number of PLL output "clicks" per second is either equal to the expected number of steps of $\pm 2\pi$ in $\psi(t)$ per second (the expected number of "clicks" per second at the output of a limiter-discriminator) or equal to zero depending on whether or not (18) is satisfied. Hence one may write

$$N_{\pm} = \frac{\gamma}{\pi} \left[\text{cerf} \left(\frac{A}{\sqrt{N}} \right) \right] U \left[\frac{\omega_L}{2\gamma} - 0.52 \right] \quad (19)$$

where $U(X)$ is the unit step function.

For the case where the limiter - PLL combination is preceded by a 13kHz rectangular band-pass filter (19) takes the form

$$N_{\pm} = \frac{BW}{\sqrt{3}} \left[\text{cerf} \left(\frac{A}{\sqrt{N}} \right) \right] U \left[\frac{2f_L}{BW} - 0.6 \right] \quad (20)$$

where $BW = 13\text{kHz}$. The expected number of PLL output "clicks" per second vs. $\frac{2f_L}{BW}$ obtained from 20 is plotted in figure 7 for $20 \log \frac{A}{\sqrt{2N}}$ equal to 2db and 4db. Corresponding experimental data are also plotted on the same set of coordinates. Although the agreement between the theoretical curves and the experimental data is not as good as it was when no limiter was present, the general shape of the curves do agree. (It would have been quite unusual to find that the experimental curves followed the sharp corner exhibited by the theoretical curves). Consequently, use of the noise model of figure 2 seems like a perfectly plausible way of determining the expected number of cycle slipping events per second in a PLL with or without a limiter.

FOOTNOTES

- (1) S. O. Rice, "Noise in FM Receivers," Chap. 25, pp. 395-424, in "Proceedings Symposium of Time Series Analysis", M. Rosenblatt (ed), John Wiley and Sons, 1963.

- (2) The radius of gyration γ of narrow band noise is given by

$$\gamma = \sqrt{\frac{\int_{-\infty}^{\infty} \omega^2 G_L(\omega) d\omega}{\int_{-\infty}^{\infty} G_L(\omega) d\omega}} \quad \text{where}$$

$G_L(\omega)$ is the low pass equivalent power spectrum of the narrow band noise centered about ω_o ⁽³⁾.

- (3) A. Papoulis, "Probability, Random Variables, and Stochastic Processes", Ch. 10, pp. 37-378, McGraw Hill, 1965.
- (4) Rice, op cit.
- (5) S. O. Rice, "Statistical Properties of a Sine Wave Plus Random Noise", Bell System Tech. J., Vol. 27, January 1948.

VII Conclusion

A model for the input carrier plus narrow-band noise to a PLL has been introduced and with its aid the expected number of PLL output "clicks" per second for a first order PLL has been calculated. This has been done for a first order loop both with and without a limiter, and in both cases the theoretical results agree quite closely with experimental data indicating the validity of the model.

Although in this report the model is applied only to a first order PLL, there is reason to believe that this model will generate fruitful results when applied to higher order PLL's or any other type of threshold extending device, such as the FMFB, or the PLL with a non-sinusoidal phase detector. In such applications, the model should lead not only to computational simplicity, but also to a better understanding of physical mechanism by which these devices provide threshold extension. Both theoretical and experimental work is underway to extend the approach in these areas.

Acknowledgment

The author is grateful to Mr. Edwin Hoffman for his independent verification of the experimental data.

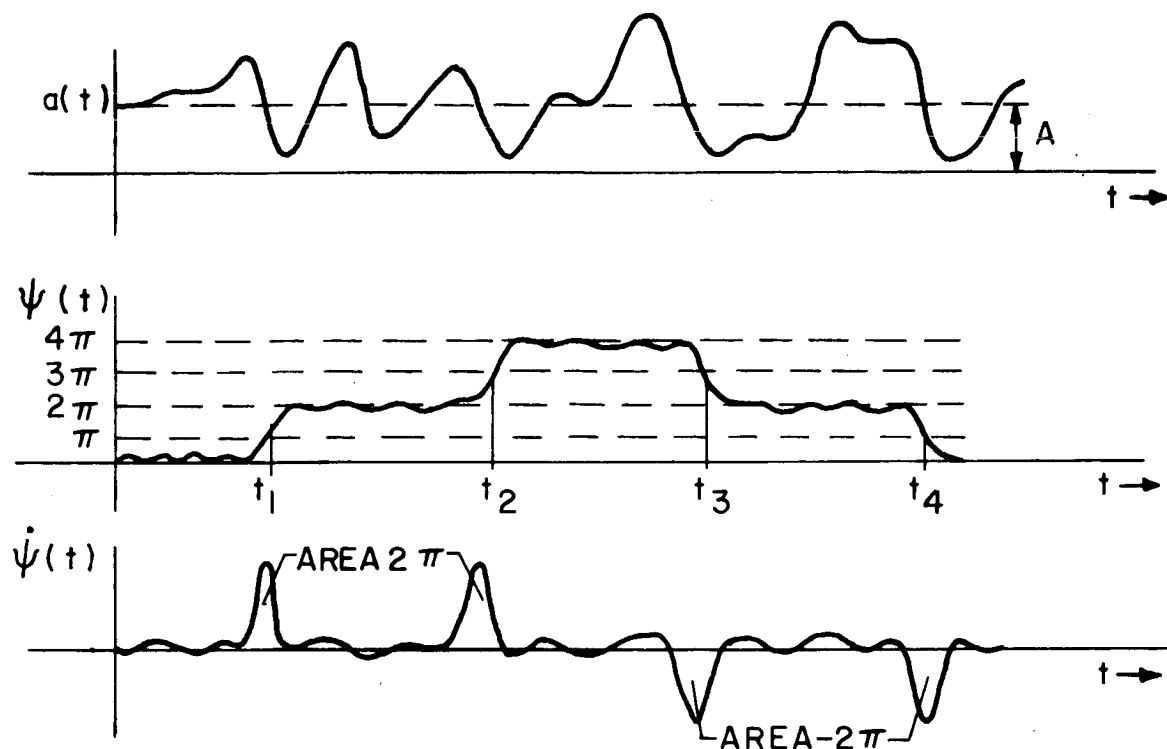


Figure 1

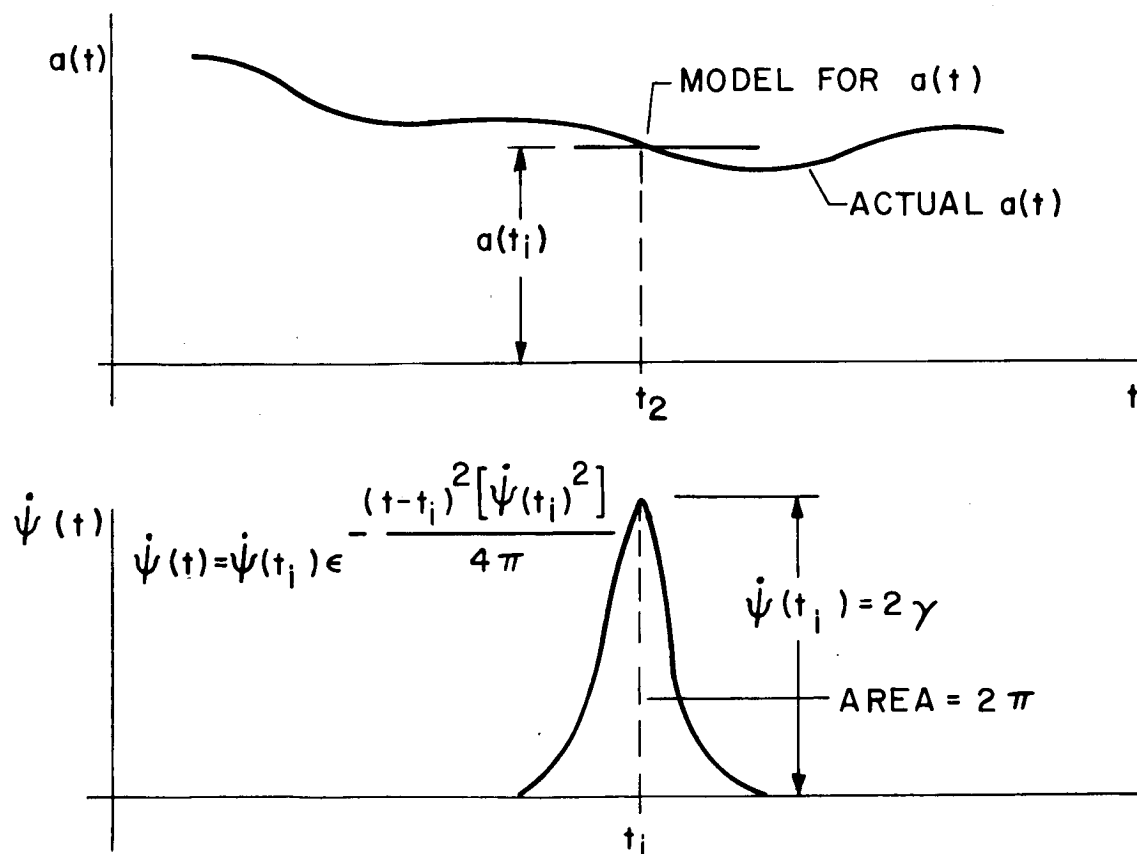


Figure 2

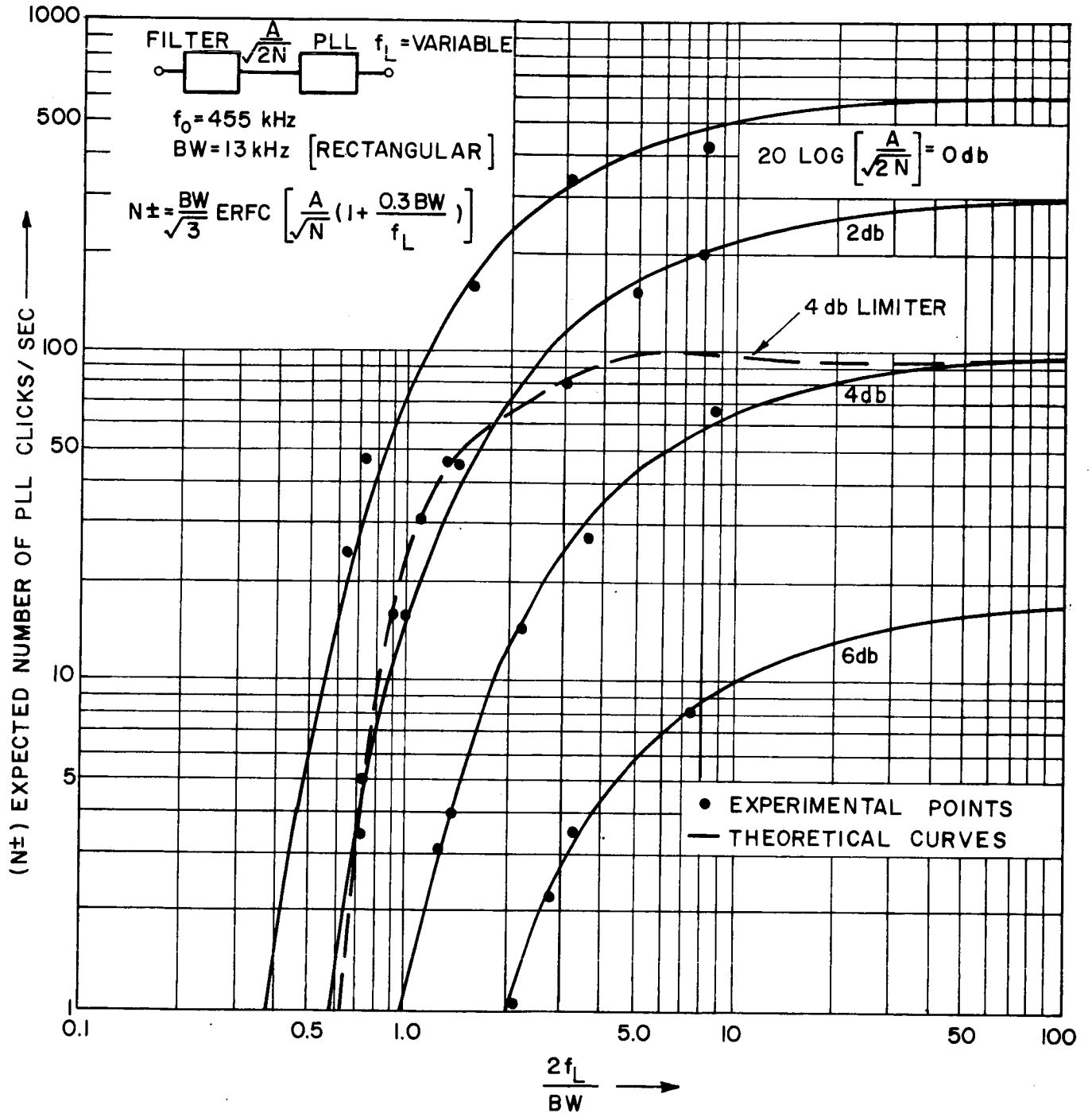


Figure 6

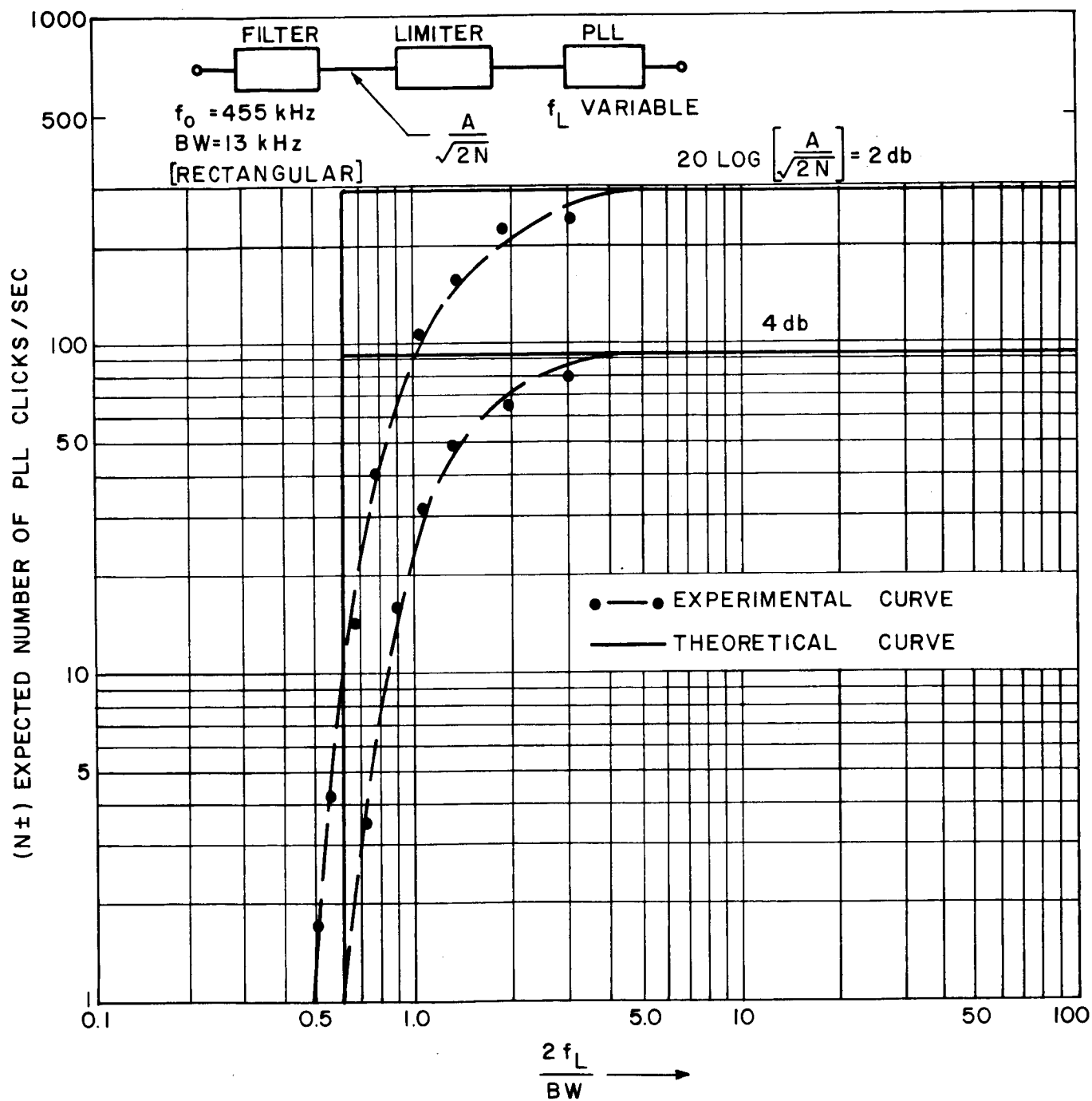


Figure 7

I-B. 2 FOKKER-PLANCK TECHNIQUES: Transient Solutions for the phase locked loop.

1. Introduction

Analyses of the phase-locked loop in the presence of noise have been concerned largely with steady-state behavior due to the difficulty of obtaining closed form solutions for the associated Fokker-Planck equation. This report summarizes the results of an investigation into the transient behavior of the first-order phase-locked loop. Numerical integration is used to obtain solutions for the Fokker-Planck equation which represents loop dynamics.

After introducing several appropriately defined constants and a dimensionless time variable, the Fokker-Planck equation is rewritten in dimensionless form. The equation is then shown to be equivalent to a one-dimensional heat flow equation in which thermal conductivity and specific heat capacity vary with position, and with a distributed heat production rate which varies with position and temperature. Expressions which establish relationships between coefficients in the two equations are derived. Numerical integration procedures which were developed by J. von Neumann and R. Richtmyer for obtaining solutions to the heat flow equation are then applied to the solution of the Fokker-Planck equation.

Results are presented for the Fokker-Planck equation which illustrate qualitative and quantitative transient behavior of the probability density function for phase error. The density function is computed for several selected values of signal-to-noise ratio for the case where the input signal frequency is known.

2. First-Order Phase-Locked Loop

2.1 Model of the Loop

A block diagram of the phase-locked loop (PLL) is shown in Figure 1. The loop input, $e_i(t)$, consists of a constant frequency sinusoidal signal plus additive white Gaussian noise,

$$e_i(t) = K_1 \sin \phi_1(t) + n(t) \quad (1)$$

where

$$\phi_1(t) = \text{input phase (rad)}$$

$$K_1 = \text{input amplitude (volts)}$$

$$n(t) = \text{white Gaussian input noise with zero mean and spectral density } N_o/2 \text{ (volts)}$$

The narrowband filter output, $e'_i(t)$, is

$$e'_i(t) = K_1 \sin \phi_1(t) + n'(t) \quad (2)$$

The signal portion of the loop input is unaffected by the filter. The additive noise, $n'(t)$, is now narrowband and, since the filter is linear, $n'(t)$ is stationary and Gaussian with zero mean. Using the narrowband representation for $n'(t)$ as derived, e. g., by Schwartz [1], section 7-9,

$$n'(t) = n_1(t) \cos \omega_o t - n_2(t) \sin \omega_o t \quad (3)$$

where

$$\omega_o = \text{center frequency of narrowband filter (rad/sec)}$$

and $n_1(t)$ and $n_2(t)$ are independent and Gaussian with zero mean. Substituting equation 3 into 2,

$$e'_i(t) = K_1 \sin \phi_1(t) + n_1(t) \cos \omega_o t - n_2(t) \sin \omega_o t \quad (4)$$

The voltage controlled oscillator (VCO) output, $e_o(t)$, is a sinusoid,

$$e_o(t) = K_2 \cos \phi_2(t) \quad (5)$$

where

$\phi_2(t)$ = output phase (rad)

K_2 = output amplitude (volts)

The frequency of the VCO output signal is controlled by the input voltage, $e(t)$, in the following manner:

$$\frac{d\phi_2(t)}{dt} = \dot{\phi}_2(t) = \omega_0 + K_3 e(t) \quad (6)$$

where

$\dot{\phi}_2(t)$ = output frequency (rad/sec)

ω_0 = VCO quiescent frequency (rad/sec)

K_3 = constant (rad/sec/volt)

In the absence of any input signal, the VCO will oscillate at constant frequency ω_0 which is called the quiescent frequency. When an input is applied, the output frequency will deviate from ω_0 by an amount proportional to the input signal.

For convenience, the input and output phase will now be defined relative to the VCO quiescent frequency,

$$\phi_1(t) = \omega_0 t + \phi'_1(t) \quad (7)$$

$$\phi_2(t) = \omega_0 t + \phi'_2(t) \quad (8)$$

where $\phi'_1(t)$ and $\phi'_2(t)$ represent relative phase.

Equations 4, 5, and 6 become, respectively,

$$e'_1(t) = K_1 \sin [\omega_0 t + \phi'_1(t)] + n_1(t) \cos \omega_0 t - n_2(t) \sin \omega_0 t \quad (9)$$

$$e_o(t) = K_2 \cos [\omega_0 t + \phi'_2(t)] \quad (10)$$

$$\dot{\phi}'_2(t) = K_3 e(t) \quad (11)$$

Equation 11 describes the dynamic behavior of the loop.

The order of a PLL is related to the number of finite poles in the filter transfer function, $K_0 F(s)$. The first-order loop, which is to be considered here, is filterless, therefore,

$$K_0 F(s) = 1 \quad (12)$$

and

$$e(t) = x(t) \quad (13)$$

The multiplier output, $x(t)$, is

$$x(t) = K_4 e_1'(t) e_0(t) \quad (14)$$

where

$$K_4 = \text{constant (volts}^{-1}\text{)}.$$

Substituting equations 9 and 10 into 14 and expanding the product terms into sum and difference frequency terms, we get

$$\begin{aligned} x(t) = & \frac{K_1 K_2 K_4}{2} \{ \sin [2\omega_0 t + \phi_1'(t) + \phi_2'(t)] \\ & + \sin [\phi_1'(t) - \phi_2'(t)] \} \\ & + \frac{K_2 K_4 n_1(t)}{2} \{ \cos [2\omega_0 t + \phi_2'(t)] + \cos \phi_2'(t) \} \\ & - \frac{K_2 K_4 n_2(t)}{2} \{ \sin [2\omega_0 t + \phi_2'(t)] - \sin \phi_2'(t) \} \end{aligned} \quad (15)$$

The VCO will not respond to the high frequency terms, therefore, they can be neglected in the VCO input,

$$\begin{aligned} e(t) = & \frac{K_1 K_2 K_4}{2} \sin [\phi_1'(t) - \phi_2'(t)] \\ & + \frac{K_2 K_4}{2} [n_1(t) \cos \phi_2'(t) + n_2(t) \sin \phi_2'(t)] \end{aligned} \quad (16)$$

The difference between input and output phase is the phase error, $\phi(t)$,

$$\phi(t) = \phi_1(t) - \phi_2(t) = \phi_1'(t) - \phi_2'(t) \quad (17)$$

The noise terms in equation 16 will be combined into a single term, $n''(t)$, whose distribution is yet to be determined,

$$n''(t) = n_1(t) \cos \phi_2'(t) + n_2(t) \sin \phi_2'(t) \quad (18)$$

Substituting equations 17 and 18 into 16,

$$e(t) = \frac{K_2 K_4}{2} [K_1 \sin \phi(t) + n''(t)] \quad (19)$$

Now, substituting equations 17 and 19 into 11,

$$\dot{\phi}(t) = \dot{\phi}_1'(t) - \frac{K_2 K_3 K_4}{2} [K_1 \sin \phi(t) + n''(t)] \quad (20)$$

The loop input, $e_1(t)$, is a constant frequency sinusoid, therefore,

$$\phi_1(t) = \omega t + \phi_{10} \quad (21)$$

where

ω = constant input frequency (rad/sec)

ϕ_{10} = initial value of the input phase (rad)

From equation 7,

$$\phi_1'(t) = (\omega - \omega_0)t + \phi_{10} \quad (22)$$

Substituting equation 22 into 20,

$$\dot{\phi}(t) = (\omega - \omega_0) - K [K_1 \sin \phi(t) + n''(t)] \quad (23)$$

where

$$K = \frac{K_2 K_3 K_4}{2} = \text{total loop gain (rad/sec/volt)} \quad (24)$$

Equations 17, 20, and 24 can be combined and represented by the loop model shown in Figure 2. This model appeared in a paper by Develet [4].

2.2 Loop Equation

Equation 23 can be rearranged and written in the form

$$\frac{d\phi(t)}{dt} - [(\omega - \omega_0) - KK_1 \sin \phi(t)] = \frac{dw(t)}{dt} \quad (25)$$

where a new process, $w(t)$, which is defined by

$$\frac{dw(t)}{dt} = -Kn''(t) \quad (26)$$

has been introduced. Since

$$dw(t) = -Kn''(t) dt \quad (27)$$

The density function, $p(\phi, t)$, of the process which satisfies differential equation 25 must satisfy the Fokker-Planck equation when the driver, $\frac{dw}{dt}$, is of the form described above (Papoulis [2], page 540), i. e. ,

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial \phi} [\eta(\phi, t)p] - \frac{1}{2} \frac{\partial^2}{\partial \phi^2} [\sigma^2(\phi, t)p] = 0 \quad (28)$$

where

$$\eta(\phi, t) dt = E[d\phi | \phi] = [(\omega - \omega_0) - KK_1 \sin \phi] dt \quad (29)$$

$$\sigma^2(\phi, t) dt = E[|dw|^2 | \phi] = \frac{K^2 N_0}{2} dt \quad (30)$$

$$p(\phi, 0) = \delta(\phi - \phi_0) \quad (31)$$

and

$$\int_{-\infty}^{\infty} p(X, t) dX = 1 \quad (32)$$

Substituting equations 29 and 30 into 31,

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial \phi} [(KK_1 \sin \phi + \omega_0 - \omega)p] + \frac{K^2 N_0}{4} \frac{\partial^2 p}{\partial \phi^2} \quad (33)$$

Equation 33 is the solution of the stochastic differential equation 23 in the sense that it defines the density function of the Markov process $\phi(t)$. It is the Fokker-Planck equation which represents the dynamic behavior of the phase-locked loop. The time varying probability density function of the loop phase error is completely characterized by equations 31, 32, and 33.

3. Fokker-Planck Equation

3.1 General

A closed form solution for equation 33 is not available. The equation has been treated by various approximations, however, there is very little quantitative information available concerning the transient behavior of the loop. Much of the work done to date assumes a stationary steady-state solution such that

$$\lim_{t \rightarrow \infty} p(\phi, t) = p(\phi) \quad (34)$$

$$\lim_{t \rightarrow \infty} \frac{\partial p(\phi, t)}{\partial t} = 0 \quad (35)$$

By substituting equation 35 into 33, an ordinary differential equation results for which a solution can be obtained modulo 2π [3], [4]. The result for the case where $\omega = \omega_0$ is

$$P(\phi) = \frac{\frac{4K_1}{KN_0} \cos \phi}{2\pi I_0(4K_1/KN_0)} e, \quad -\pi \leq \phi < \pi \quad (36)$$

where

$$P(\phi, t) = \sum_{k=-\infty}^{k=\infty} p(\phi + 2\pi k, t) \quad (37)$$

and I_0 is a modified Bessel function of zero order. The reason for taking ϕ modulo 2π is that if a steady state solution exists,

$$\lim_{t \rightarrow \infty} p(\phi, t) = p(\phi) = 0, \quad -\infty \leq \phi < \infty \quad (38)$$

However, by "stacking" segments of $p(\phi)$ which are of length 2π into the interval $-\pi \leq \phi < \pi$, the result given by equation 36 is obtained.

Another useful approach is to consider only small phase errors such that

$$\sin \phi \approx \phi \quad (39)$$

By substituting equation 39 into 33, a linearized version of the loop equation is obtained. It will be convenient at this point to express equation 33 in dimensionless form. Let

$$B_L = \frac{KK_1}{4} = \text{loop bandwidth (rad/sec)} \quad (40)$$

$$\alpha = \frac{K_1^2}{N_o B_L} = \text{signal-to-noise ratio (SNR) in the bandwidth of the loop (dimensionless)} \quad (41)$$

$$\gamma = \frac{\omega - \omega_o}{4 B_L} = \text{frequency ratio (dimensionless)} \quad (42)$$

$$\tau = 4 B_L t = \text{dimensionless time} \quad (43)$$

Substituting equations 40 through 43 into 33 using the relation

$$\frac{\partial p}{\partial t} = \frac{\partial p}{\partial \tau} \frac{\partial \tau}{\partial t} = 4 B_L \frac{\partial p}{\partial \tau} \quad (44)$$

we get the dimensionless Fokker-Planck loop equation

$$\frac{\partial p}{\partial \tau} = \frac{\partial}{\partial \phi} [(\sin \phi - \gamma) p] + \frac{1}{\alpha} \frac{\partial^2 p}{\partial \phi^2} \quad (45)$$

Considering again the case where $\omega = \omega_o$ and substituting equation 39, the linearizing approximation, results in the following:

$$\frac{\partial p}{\partial \tau} = \frac{\partial}{\partial \phi} (\phi p) + \frac{1}{\alpha} \frac{\partial^2 p}{\partial \phi^2} \quad (46)$$

The closed form solution of equation 46, which satisfies equations 31 and 32, is

$$p(\phi, \tau) = \sqrt{\frac{\alpha}{2\pi(1 - e^{-2\tau})}} e^{-\frac{\alpha(\phi - \phi_o e^{-\tau})^2}{2(1 - e^{-2\tau})}} \quad (47)$$

The solution for the case where $\omega \neq \omega_0$, is,

$$p(\phi, \tau) = \sqrt{\frac{\alpha}{2\pi(1 - e^{-2\tau})}} e^{-\frac{\alpha[\phi - \gamma - (\phi_0 - \gamma)e^{-\tau}]^2}{2(1 - e^{-2\tau})}} \quad (48)$$

Note that $p(\phi, \tau)$ is Gaussian with mean, m , and variance, σ^2 , given by

$$m(\tau) = \gamma + (\phi_0 - \gamma)e^{-\tau} \quad (49)$$

$$\sigma^2(\tau) = \frac{1 - e^{-2\tau}}{\alpha} \quad (50)$$

to satisfy the restriction on Z . There is no reason to expect singularities in p (which would result from an unstable equation) other than at $p(\phi, 0)$. In any event, if singularities do exist, they will become evident when the calculating routine fails to converge for various values of p . As we will see, this does not occur.

3.2 Numerical Solution

Numerical solutions for the Fokker-Planck loop equation were obtained using a calculating routine established in [5] for the solution, the heat flow equation.

The partial difference system which was used can best be described by assuming a rectangular mesh of points (ϕ_ℓ, τ_n) in the $\phi - \tau$ plane as illustrated in Figure 3. The value of a function, e.g., $p(\phi, \tau)$, at mesh point (ϕ_ℓ, τ_n) will be denoted by p_ℓ^n . First derivatives — consider, for example, $\frac{\partial p}{\partial \phi}$ — are approximated as follows,

$$\begin{aligned} \frac{\partial p}{\partial \phi} &\approx \frac{p(\phi_\ell + \Delta\phi, \tau_n) - p(\phi_\ell, \tau_n)}{\Delta\phi} \\ &= \frac{p_{\ell+1}^n - p_\ell^n}{\Delta\phi} \end{aligned} \quad (51)$$

or

$$\begin{aligned}\frac{\partial p}{\partial \phi} &\approx \frac{p(\phi_\ell, \tau_n) - p(\phi_\ell - \Delta\phi, \tau_n)}{\Delta\phi} \\ &= \frac{p_\ell^n - p_{\ell-1}^n}{\Delta\phi}\end{aligned}\tag{52}$$

where equations 57 and 52 are the forward and backward first-difference quotients, respectively. Both difference quotients approximate the value of $\frac{\partial p}{\partial \phi}$ at mesh point (ϕ_ℓ, τ_n) . Note in Figure 3 that $\Delta\tau$ and $\Delta\phi$ are the step sizes in τ and ϕ .

Second derivatives are approximated as follows,

$$\frac{\partial}{\partial \phi} \left(\frac{\partial p}{\partial \phi} \right) \approx \frac{\frac{p_{\ell+1}^n - p_\ell^n}{\Delta\phi} - \frac{p_\ell^n - p_{\ell-1}^n}{\Delta\phi}}{\Delta\phi}\tag{53}$$

To approximate a term such as

$$\frac{\partial}{\partial \phi} \left(Y \frac{\partial p}{\partial \phi} \right)$$

we will use the values of Y midway between (ϕ_ℓ, τ_n) and $(\phi_{\ell+1}, \tau_n)$ which we denote $Y_{\ell+1/2}^n$ and $Y_{\ell-1/2}^n$. The partial difference system which we will utilize approximates $Y_{\ell+1/2}^n$ by the mean of Y_ℓ^n and $Y_{\ell+1}^n$, i. e. ,

$$Y_{\ell+1/2}^n = \frac{Y_\ell^n + Y_{\ell+1}^n}{2}\tag{54}$$

$$Y_{\ell-1/2}^n = \frac{Y_{\ell}^n + Y_{\ell-1}^n}{2} \quad (55)$$

Then, using equation 86,

$$\frac{\partial}{\partial \phi} \left(Y \frac{\partial p}{\partial \phi} \right) \approx \frac{Y_{\ell+1/2}^n (p_{\ell+1}^n - p_{\ell}^n) - Y_{\ell-1/2}^n (p_{\ell}^n - p_{\ell-1}^n)}{(\Delta \phi)^2} \quad (56)$$

We can now write the partial difference system which will be used to approximate equation 68,

$$\frac{p_{\ell}^{n+1} - p_{\ell}^n}{\Delta \tau} = \frac{X_{\ell}^n \left[Y_{\ell+1/2}^n (p_{\ell+1}^n - p_{\ell}^n) - Y_{\ell-1/2}^n (p_{\ell}^n - p_{\ell-1}^n) \right]}{(\Delta \phi)^2} + Z_{\ell}^n \quad (57)$$

or, solving for p_{ℓ}^{n+1} , the value of p at time τ_{n+1} , in terms of values of the various functions at time τ_n ,

$$p_{\ell}^{n+1} = p_{\ell}^n + \frac{\Delta \tau X_{\ell}^n \left[Y_{\ell+1/2}^n (p_{\ell+1}^n - p_{\ell}^n) - Y_{\ell-1/2}^n (p_{\ell}^n - p_{\ell-1}^n) \right]}{(\Delta \phi)^2} + \Delta \tau Z_{\ell}^n \quad (58)$$

Note that the forward first-difference quotient is used to approximate $\frac{\partial p}{\partial \tau}$ since we wish to step forward in time

Equation 58 is an explicit difference system in that the result can be computed step by step. Knowing p_{ℓ}^n we compute p_{ℓ}^{n+1} , then p_{ℓ}^{n+2} , etc. An alternative approach utilizes an implicit difference system which relates the p_{ℓ}^{n+j} by a set of simultaneous equations, the solution of which requires inversion of a very large matrix.

3.3 Boundary Condition and Initial Condition

It is impractical to attempt to compute $p(\phi, \tau)$ over the entire range of ϕ . Therefore, it becomes necessary to select an appropriate interval of computation and associated boundary conditions. Note that the computation of p_{ℓ}^{n+1} using equation 58 requires values of $p_{\ell \pm 1}^n$, i. e., to compute the value of a point on the boundary we must use the value of a point outside the boundary. This suggests the following boundary conditions when $p(\phi, \tau)$ is computed over the interval $\phi_i \leq \phi \leq \phi_j$:

$$p_{i-1}^n \approx p_{i+1}^n \quad (59)$$

$$p_{j+2}^n \approx p_{j-1}^n \quad (60)$$

These boundary conditions assume that $p(\phi, \tau)$ is symmetrical about the boundaries for values of ϕ very near the boundaries. The approximations approach equalities as mesh size is made smaller and smaller. Further justification arises from the fact that the range over which $p(\phi, \tau)$ is computed will normally be selected sufficiently large so that the boundary values are relatively small, and can therefore introduce little error into the overall result.

The initial value of $p(\phi, \tau)$ is defined by equation 31. Since this initial value, $\delta(\phi - \phi_0)$, is unbounded in magnitude, it must be approximated by some function which satisfies equation 32. A Gaussian density function with mean ϕ_0 and very small variance, σ_0^2 , was selected as the approximating function.

Since $p(\phi, \tau)$ is a density function its area must integrate to unity.

Although we do not consider the entire range of ϕ in our numerical calculations, for small τ we would expect that

$$\int_{\phi_i}^{\phi_j} p(V, \tau) dV = 1 \quad (61)$$

where ϕ_i and ϕ_j are the boundary values introduced above. As the interval $\phi_i \leq \phi \leq \phi_j$ is expanded, equation 61 will hold true over an increasing interval of time. It becomes clear that we have an accuracy check which can be applied over an arbitrary number of time steps.

4. Results

4.1 Density Function of Phase Error

Equation 58, the partial difference system, was programmed for computation on a digital computer. Results were obtained for several values of α with

$$\phi_0 = 0 \quad (62)$$

$$\gamma = \omega - \omega_0 = 0 \quad (63)$$

The value used for σ_0 , the standard deviation of the initial Gaussian density function, was $\pi/20$. Results were obtained over the interval $-3\pi \leq \phi \leq 3\pi$, however, since $p(\phi, \tau)$ is an even function for the conditions stated in equations 62 and 63, it was necessary to compute $p(\phi, \tau)$ only over the interval $0 \leq \phi \leq 3\pi$.

Results obtained for $\alpha = 10$ are illustrated in Figure 4. As can be seen, only a limited range of τ was considered here. It was decided to concentrate on smaller values of α so as to obtain more rapid variation of $p(\phi, \tau)$ with time.

Results for $\alpha = 1$ appear in Figure 5, which illustrates the behavior of $p(\phi, \tau)$ in the interval $-\pi \leq \phi \leq \pi$. The variation of $p(\phi, \tau)$ in the range $\pi \leq |\phi| \leq 3\pi$ is seen more clearly when even smaller values of α are considered.

Shown in Figure 6 are the results obtained for $\alpha = 0.1$. We see here the variation of $p(\phi, \tau)$ for $|\phi| > \pi$. Results are plotted to a larger scale in Figure 7 for selected values of τ . Equation 47, the solution for small ϕ , was computed for $\alpha = 0.1$ and the results superimposed on Figure 7. It was determined that when $\tau = .0012$, equation 47 is identical to the initial Gaussian density function used. Therefore, equation 47 was actually computed for $\tau = \tau' + .0012$, where $\tau'/4$ (see equation 43) corresponds to the dimensionless time values indicated in Figure 7.

5. Conclusion

5.1 Summary

We see that the partial difference system given in equation 58 can be used to obtain numerical solutions for the Fokker-Planck loop equation. This approach treats the loop equation as the following one-dimensional heat flow equation

$$\frac{\partial p}{\partial \tau} = \frac{1}{\alpha} e^{\alpha (\cos \phi + \gamma \phi)} \left\{ \frac{\partial}{\partial \phi} \left[\left(e^{-\alpha (\cos \phi + \gamma \phi)} \right) \frac{\partial p}{\partial \phi} \right] \right\} + p \cos \phi \quad (64)$$

and then applies a numerical method developed by von Neumann and Richtmyer for obtaining solutions.

Although equation 64 contains periodic terms, periodicities do not appear in the numerical solutions for the range of τ considered. Transient behavior is seen to remain exponential in character. Of course, it cannot be concluded that periodic behavior will not arise for larger values of τ . In fact, the only conclusion warranted in this respect is that any periodic behavior present becomes evident only after the ordinates of $p(\phi, \tau)$ have become small.

As an additional check on the validity of results obtained, values of $p(\phi, \tau)$ for $\alpha = 1$ from Figure 5 are plotted modulo 2π in Figure 11 and compared with the steady state solution as given in equation 36. A similar result for $\alpha = .01$ appears in Figure 12.

5.2 Second-Order Loop

In deriving the Fokker-Planck equation for the first-order loop, it was assumed that the loop contained no filter (see equation 12). The second-order loop does contain a filter with transfer function of the form

$$K_C F(s) = \frac{K_C (s + a)}{s} \quad (65)$$

The Fokker-Planck equation for the second-order loop is derived by Viterbi [4], as follows,

$$\frac{\partial p}{\partial t} = -X_1 \frac{\partial p}{\partial X_0} + \frac{\partial}{\partial X_1} \left\{ \left[K K_1 \sin (a X_0 + X_1) \right] p \right\} + \frac{K^2 N_0}{4} \frac{\partial^2 p}{\partial X_1^2} \quad (66)$$

We see that the second-order loop equation is two-dimensional. If it were desired to apply the approach discussed in this report to the second-order loop, the initial step required would be to establish a partial difference system which approximates equation 66. The approach here would be similar to that discussed in Section 3.2. For example, the forward first-difference quotient analogous to equation 51 would be

$$\frac{\partial p}{\partial X_0} \approx \frac{p(X_{0\ell} + \Delta X_0, X_{1m}, t_n) - p(X_{0\ell}, X_{1m}, t_n)}{\Delta X_0} \quad (67)$$

Similar quantities analogous to equations 57 and 58 can also be written. However, as we have seen in Section 3.3, the primary difficulty lies not in establishing the partial difference system but in deriving stability conditions. A possible approach here would be an extension of von Neumann's derivation, as summarized in Appendix II, so as to obtain stability conditions for the two-dimensional case of the heat flow equation.

A possible alternate approach for investigating transient behavior of the second-order loop using numerical methods is Monte Carlo simulation. This method would involve a digital simulation of the loop driven by Gaussian white noise which is generated by weighting the output of a random number generator according to a normal distribution. An experiment consists of determining the system response for a single randomly generated noise value. Then, by repeating the experiment hundreds or thousands of times, statistical estimates of the system response (e.g., the probability density function) can be obtained.

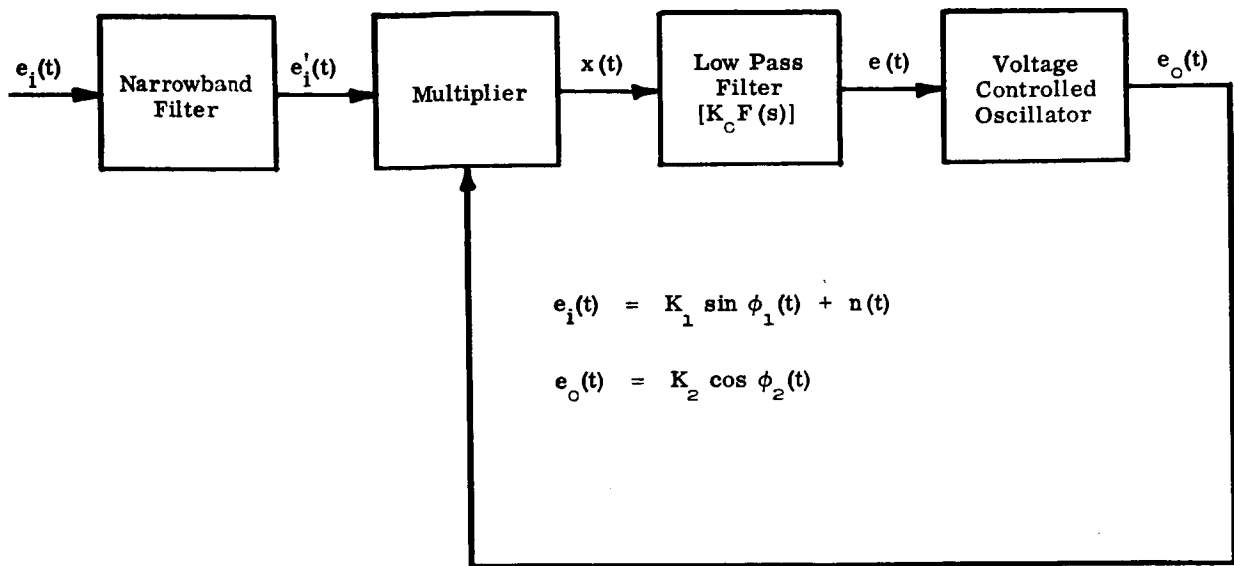
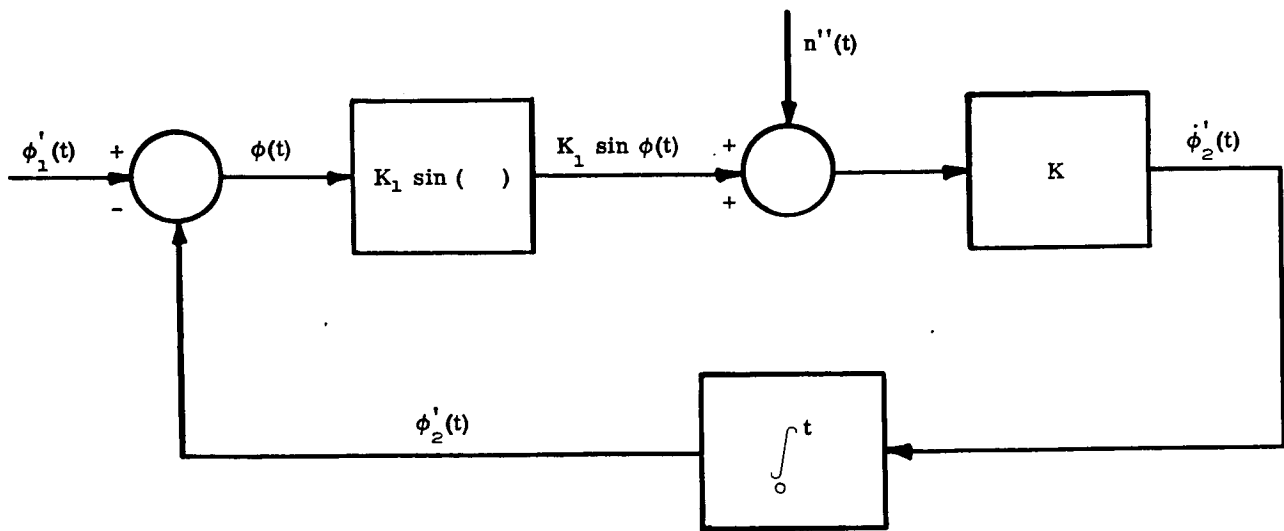


Figure 1. Phase-Locked Loop



$$\dot{\phi}(t) = (\omega - \omega_o) - K[K_1 \sin \phi(t) + n''(t)]$$

$$\phi_1'(t) = \omega - \omega_o$$

$$\phi(t) = \phi_1'(t) - \phi_2'(t)$$

Figure 2. Model of the Phase-Locked Loop

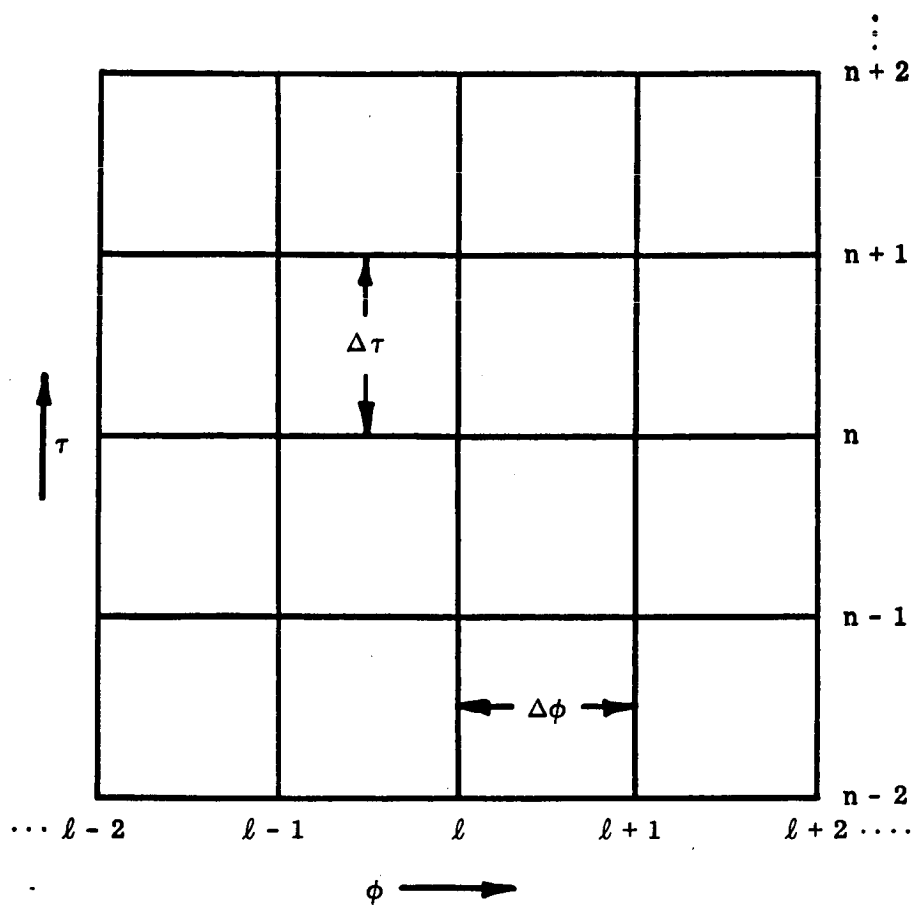


Figure 3. Rectangular Computing Grid

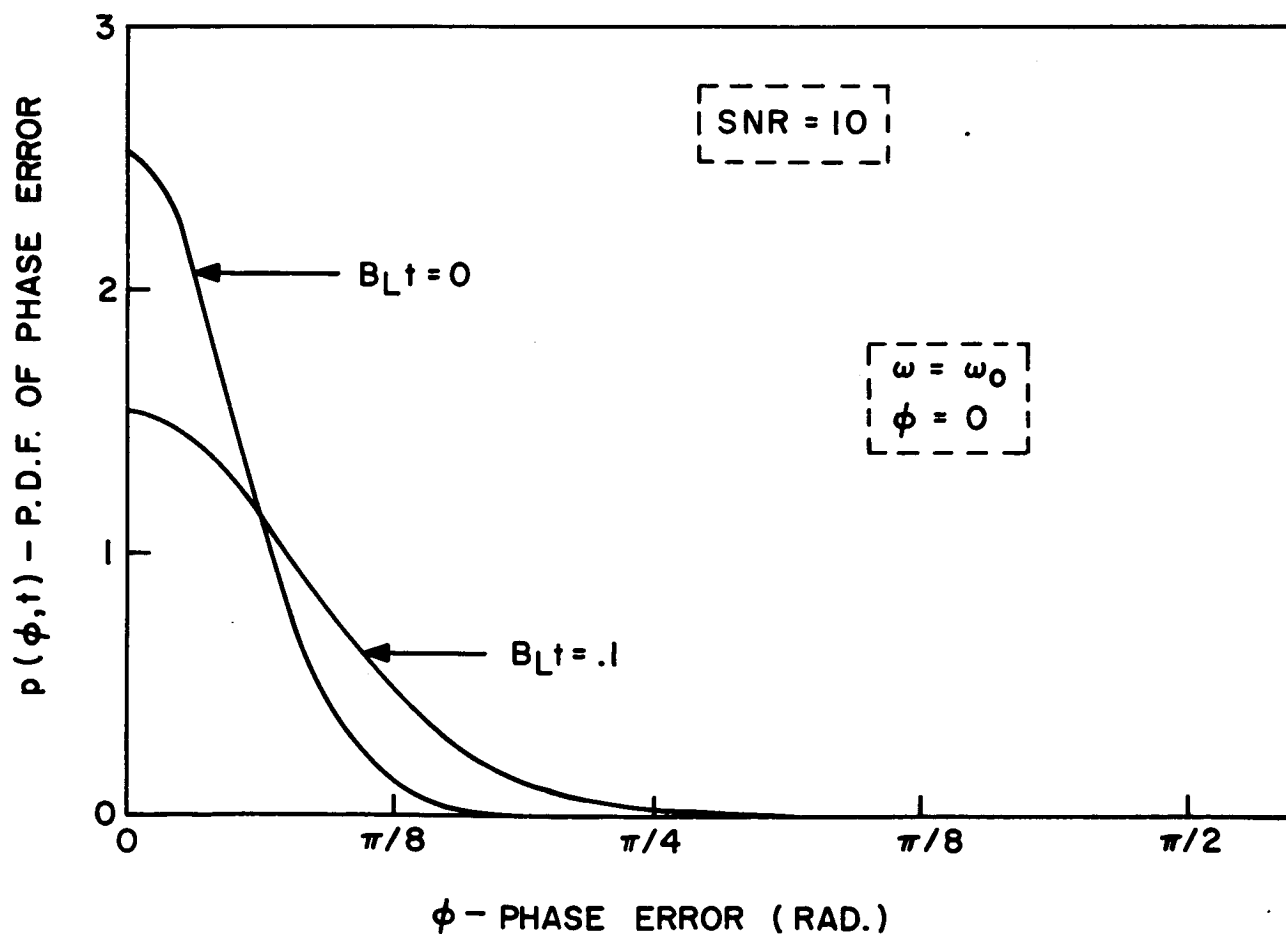


Figure 4. Probability Density Function of Phase Error

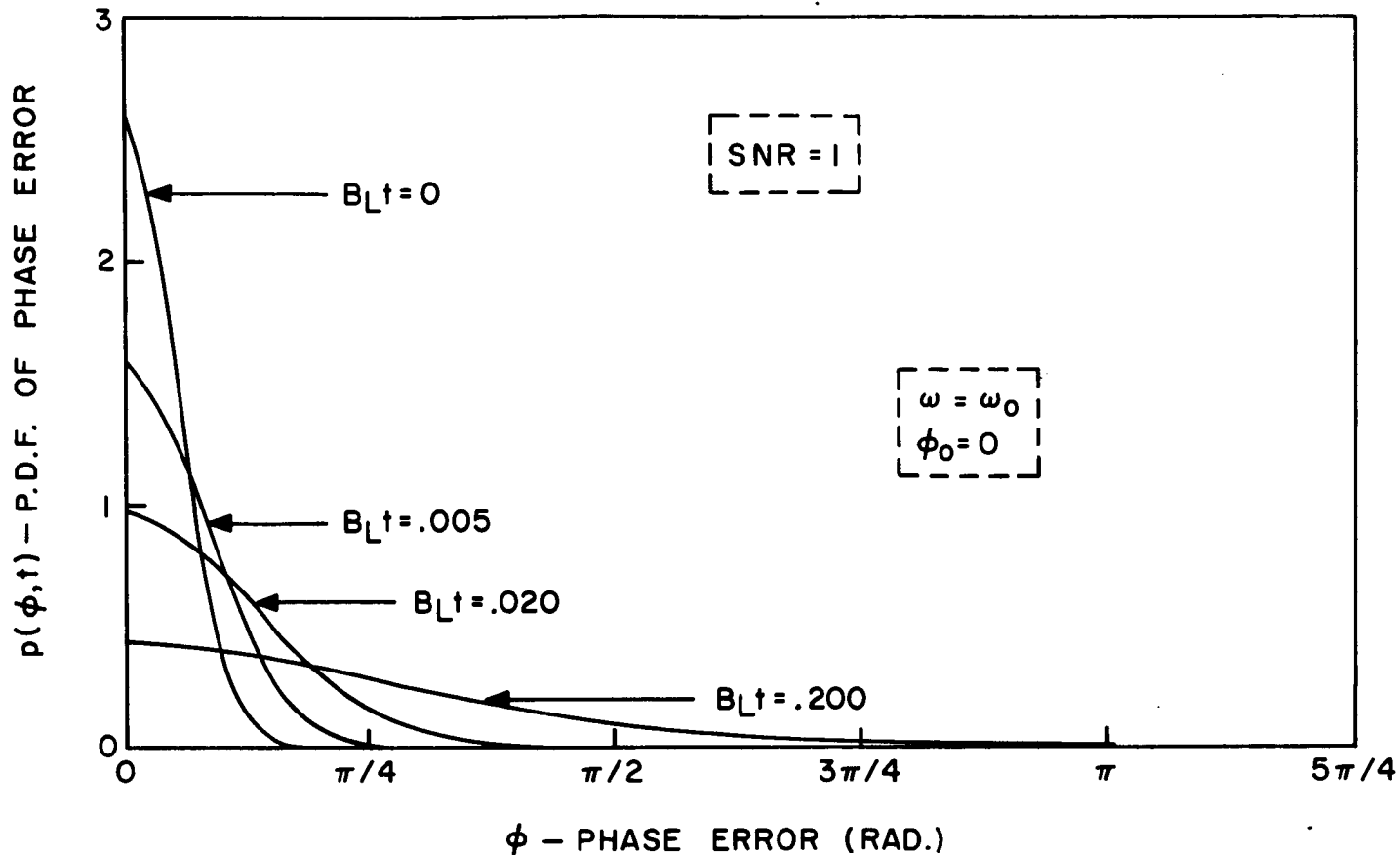


Figure 5. Probability Density Function of Phase Error

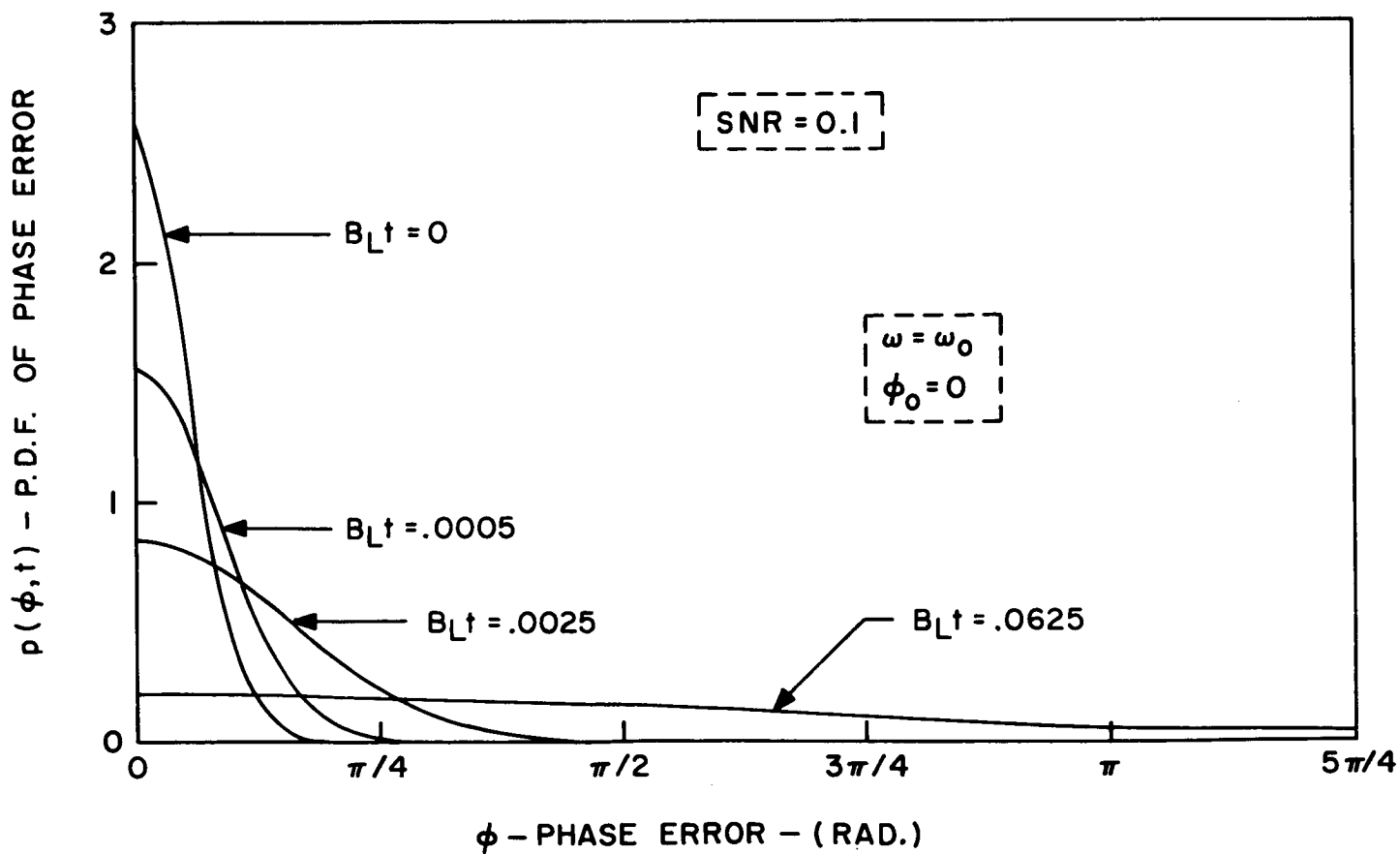


Figure 6. Probability Density Function of Phase Error

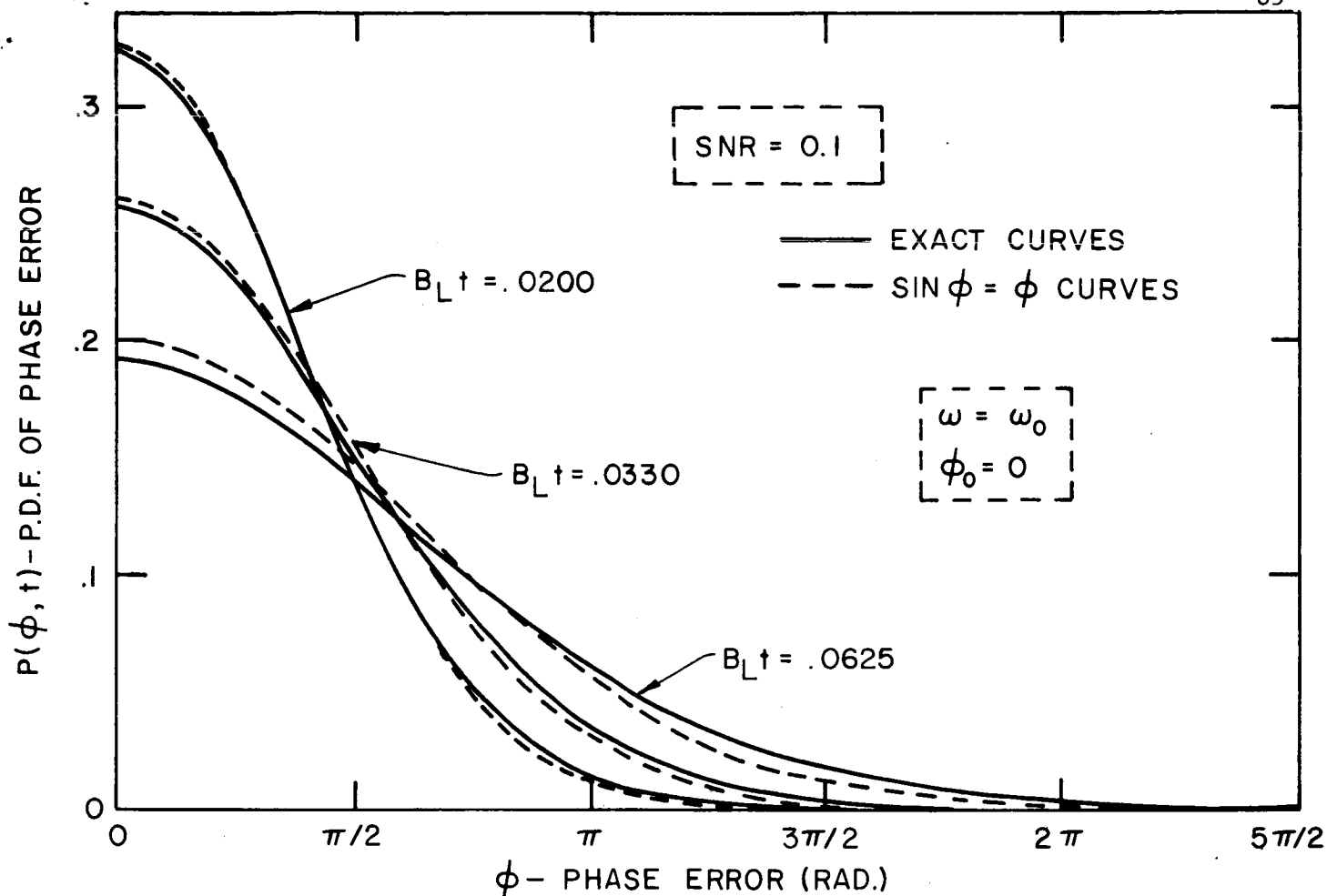


Figure 7. Probability Density Function of Phase Error

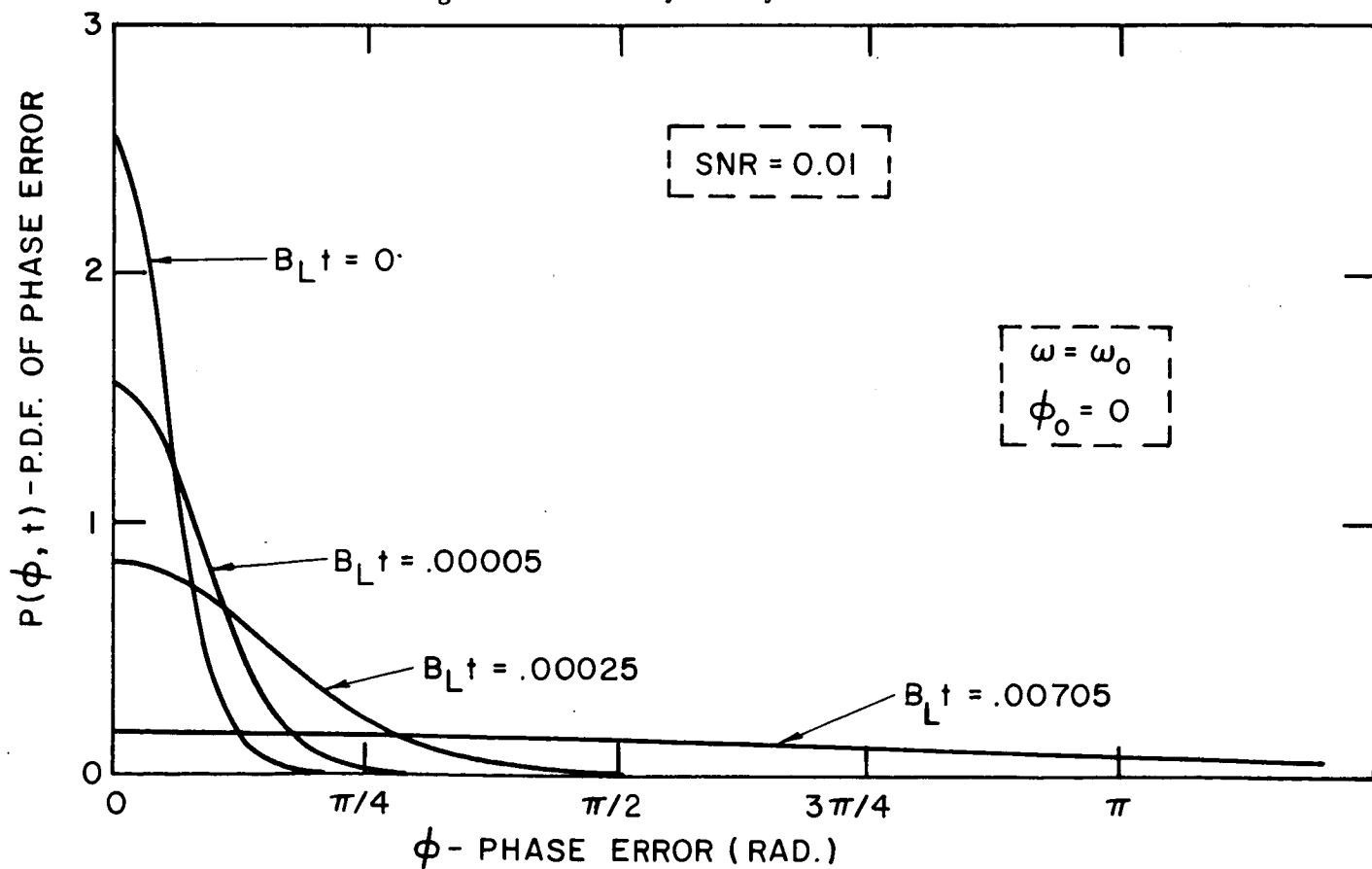


Figure 8. Probability Density Function of Phase Error

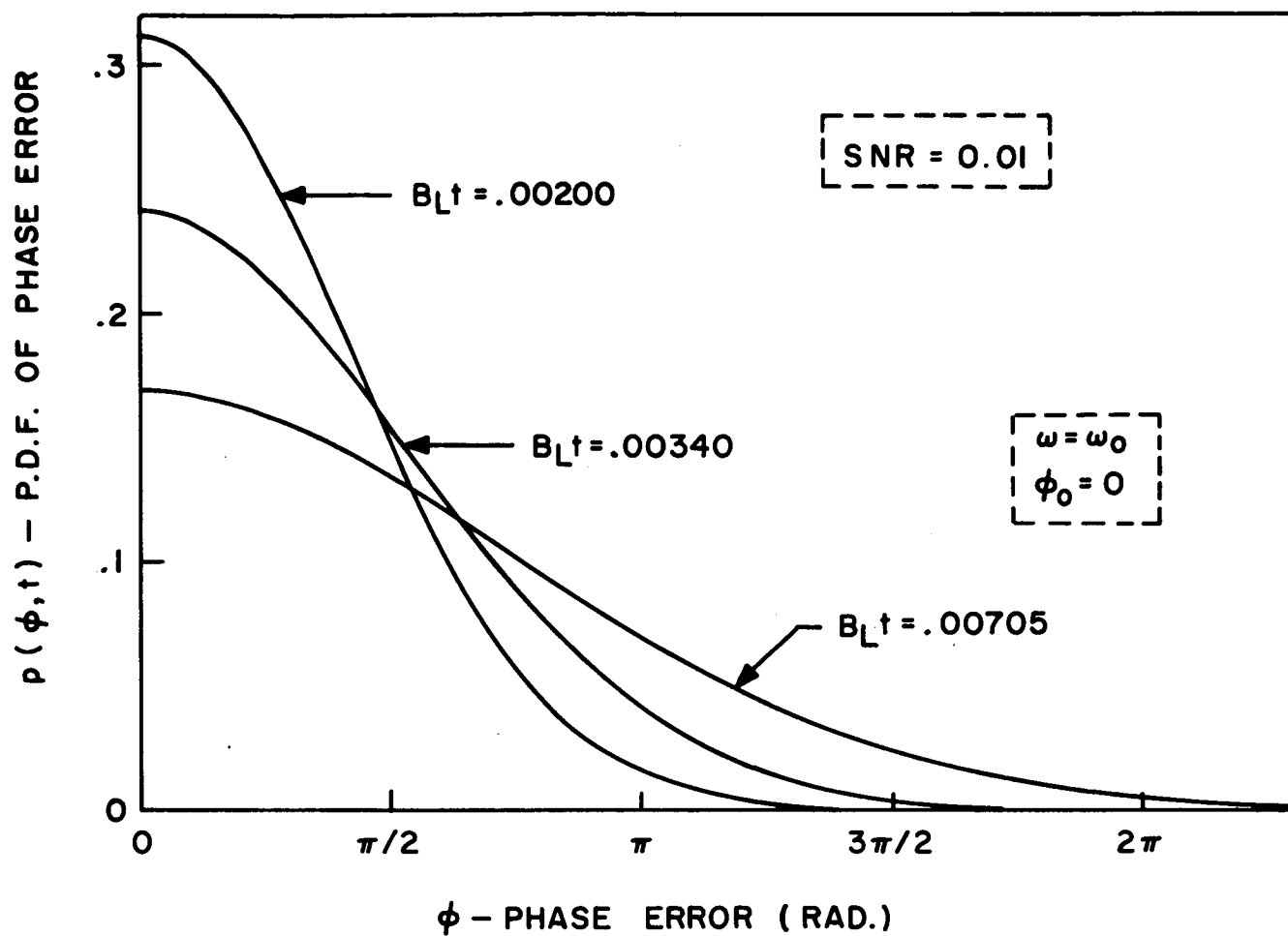


Figure 9. Probability Density Function of Phase Error

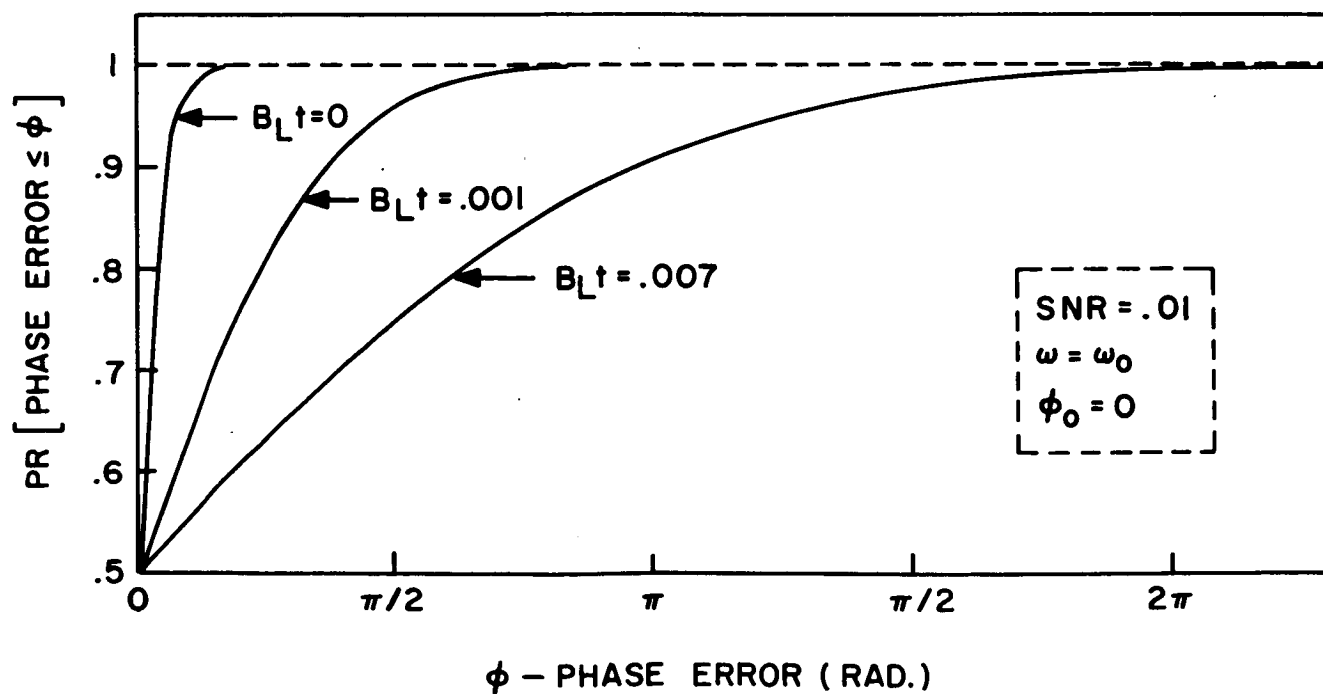


Figure 10. Probability Distribution Function of Phase Error

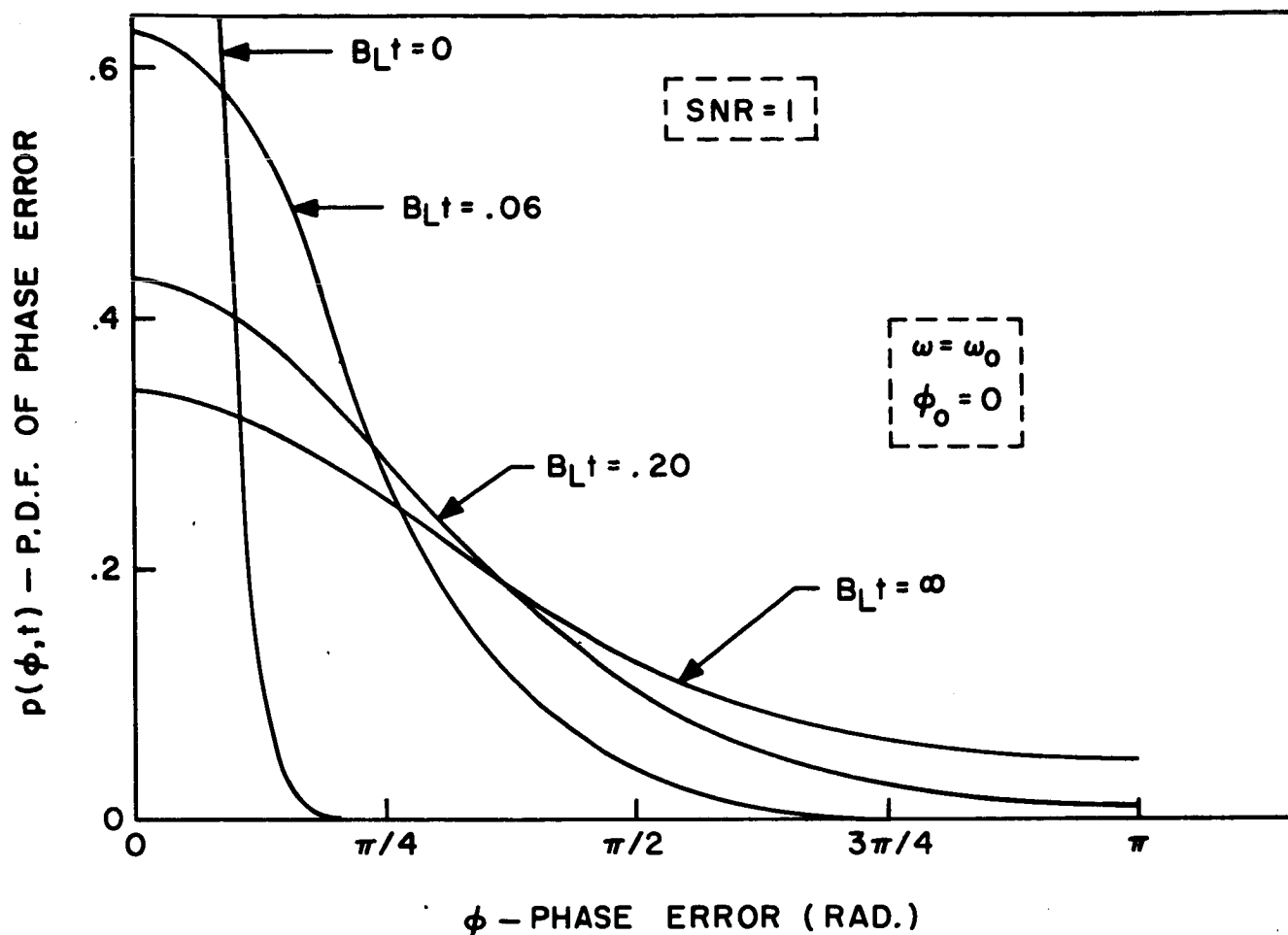


Figure 11. Probability Density Function of Phase Error with ϕ taken Modulo 2π

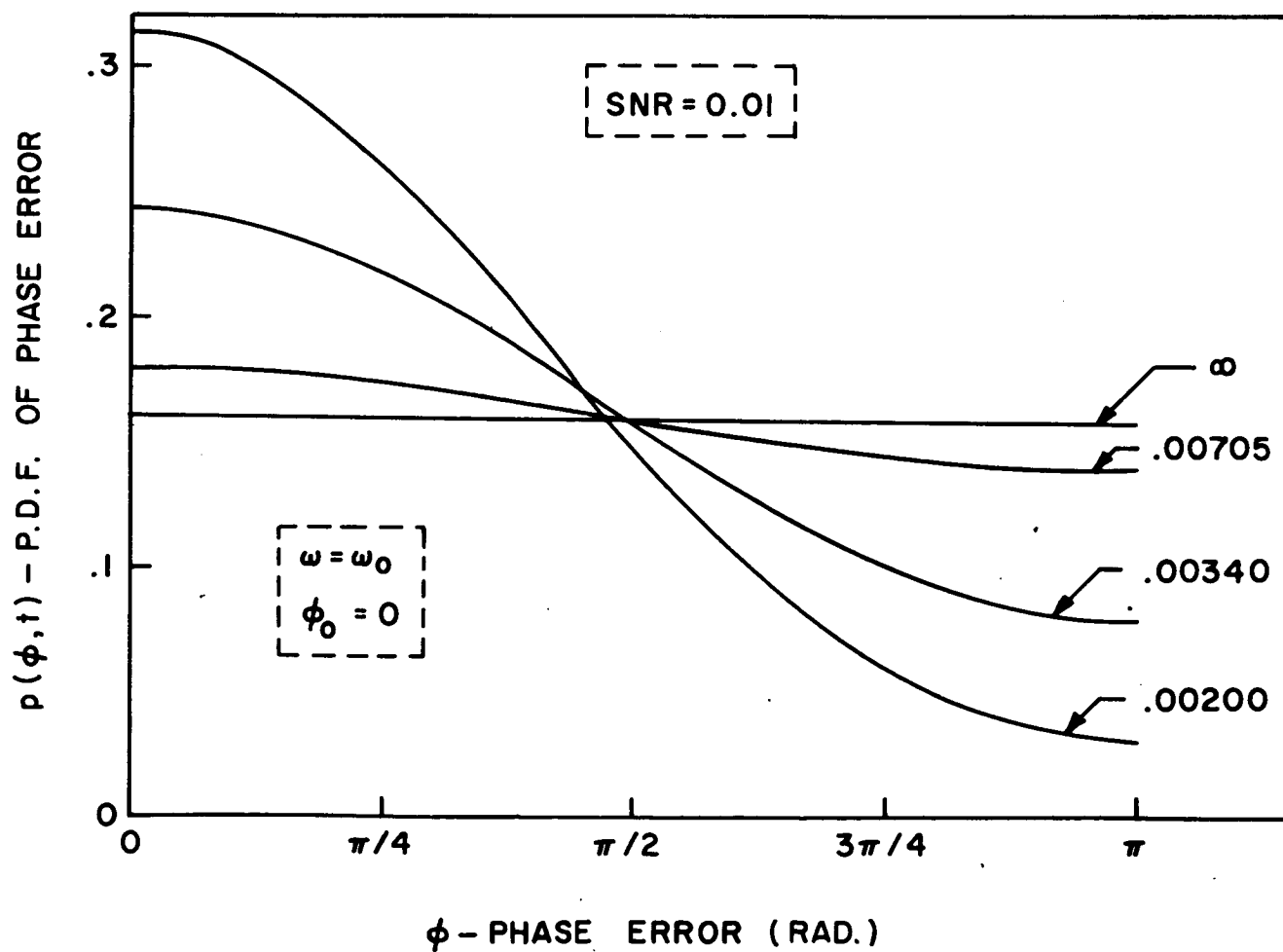


Figure 12. Probability Density Function of Phase Error with ϕ taken Modulo 2π

Bibliography

- [1] Schwartz, M., Information Transmission, Modulation, and Noise, McGraw-Hill Book Co., Inc., New York, 1959.
- [2] Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Co., Inc., New York, 1965.
- [3] Viterbi, A.J., Principles of Coherent Communication, McGraw-Hill Book Co., Inc., New York, 1966.
- [4] Develet, J. A., A Threshold Criterion for Phase-Lock Demodulation, Proc. IRE, Vol. 51, No. 2, pp. 349-356, February 1963.
- [5] von Neumann, J. and Richtmyer, R.D., On the Numerical Solution of Partial Differential Equations of Parabolic Type, Los Alamos Report LA657, December 25, 1947. Reprinted in John von Neumann-Collected Works, Volume V, Edited by A.H. Taub, Macmillan Co., New York, 1963.

I-B.3 Drift Correction for VCO in PLL

Introduction

A first-order phase-locked loop (PLL) used to demodulate a frequency modulated carrier is shown in Fig.1 . The phase comparator provides an output which depends on the phase difference between the input signal and the output of the VCO. This phase-comparator signal constitutes both the demodulated output and the control signal which constrains the VCO to lock, in frequency, to the input signal. There is a limited range of phase over which the phase comparator is able to provide a control signal adequate to achieve locking. For example, if phase comparison is accomplished by multiplication, the total range of phase is 180° and the center of the range corresponds to the situation where the two inputs to the comparator are in quadrature. The range of input frequency over which the PLL will remain in lock, i. e. its bandwidth depends on the amplitudes of the input signal and VCO output as well as on the sensitivity with which the VCO responds to its control signal.

When the input to the circuit of Fig. 1 consists of an unmodulated carrier accompanied by a high level of noise, the demodulator output displays a combination of "smooth" noise and "spike" noise. The spike noise⁽¹⁾ which has been the subject of intensive recent study is responsible for the threshold phenomenon encountered in frequency modulation communication systems.

In this case of unmodulated carrier plus noise input, it is observed experimentally that, when the VCO is adjusted so that the phase comparator is operating in the center of its range, the frequency of these randomly occurring spikes is at a minimum, and that the average rates of occurrence of positive spikes and of negative spikes are equal. When the VCO is adjusted so that the comparator is operating progressively off center the spikes of different polarities become progressively unequal in rate of occurrence. The polarity of the predominant spike reverses when the comparator moves from one side to the other side of the center of its range. Furthermore, and most importantly, the total number of the sum of positive and negative spikes increases sharply when the comparator operating point departs in either direction from center.

When the carrier is modulated, the total rate of occurrence of spikes increases over the unmodulated case since the phase comparator will be off center except when the instantaneous frequency of the input momentarily equals the carrier frequency. For the purpose of minimizing spike noise it is clearly advantageous to arrange that the phase comparator operating point hover about its center of range. If the modulation carries the instantaneous frequency symmetrically about the carrier frequency then the advantage is in adjusting the VCO so that the carrier frequency is at the center of the range of the PLL.

In any physical system, the frequency setting of the VCO may be expected to drift and such drift will degrade the performance of the system. In telemetry systems the baseband frequency range may extend to d-c. In such a case, and when high sensitivity is required in the VCO, the phase comparator may have to be coupled to the VCO through a high-gain d-c amplifier thereby compounding the drift problem.

Accordingly, work was undertaken to develop circuitry to correct for VCO drift. The circuitry is based on the recognition that, depending on the direction of the drift, spikes of one or the other polarity will predominate. A drift-correcting signal is applied to the VCO where magnitude and polarity depend on the extent and direction of the unbalance of spikes of the two polarities .

I. UNMODULATED CARRIER

(a) Circuit Block Diagram

To check the feasibility of the VCO drift-correcting proposal suggested above, the scheme was tried initially in connection with an unmodulated carrier. A block diagram of the arrangement employed is shown in Fig. 2 . Since the input carrier is unmodulated, the phase comparator output consists of noise above, smooth noise and spike noise . The noise is amplified and applied to two diode voltage comparators. The reference level of one comparator is adjusted so that it transmits only the peaks of the negative spikes while the other comparator transmits only the peaks of the positive spikes. The smooth noise, having a smaller amplitude, is not transmitted. These spike peaks are used to trigger the monostable multivibrators as shown. The multivibrator output pulses are of equal amplitude and duration, and of opposite polarity. The width of the multivibrator outputs are adjusted to be much wider than the spike width, but narrow enough to make insignificant the likelihood that two spikes of the same polarity will occur during this multivibrator pulse duration. The multivibrator pulses are much larger in amplitude than the triggering spikes.

The multivibrator outputs are added and the sum applied to a low-pass resistance-capacitance combination with a time constant which is very long in comparison to the interval between spikes. The output of this integrator is proportional to the long-time average of the difference between the numbers of spikes of the two polarities. This slowly varying signal,

after further amplification to increase the sensitivity of the feedback control, is applied as a correcting signal to the VCO.

(b) Schematic Circuit

A schematic circuit is shown in Fig. 3. The four-diode ring and its associated transformers T1 and T2 constitute the phase comparator. The stage involving Q1 is an emitter follower used as a buffer. The stage involving Q2 is an amplifier on which the transistor is biased to respond somewhat more sensitively to a positive input signal excursion than to a negative excursion. The simplified and inverted output therefore appears with an accentuated negative going spike. Diode D1 and its associated resistors constitutes the voltage comparator which allows transmission of only the spike extremity thereby preventing the triggering of the multivibrator (Q4 and Q5) by the smooth noise. The voltage comparison reference level is set by R1. The output at the collector of Q5 is a positive-going multivibrator pulse. The channel for the generation of a negative going multivibrator pulse involves transistors Q3, Q6 and Q7 and diode D2. Observe that one channel uses n-p-n transistors while the other uses p-n-p transistors.

The addition of opposite polarity multivibrator pulses and the long time averaging is accomplished with resistors R3, R4 and capacitor C1. The time constant associated with this circuit is about 7 seconds. The resistor R5 serves to maintain the voltage across C1, an electrolytic capacitor. The field-effect transistor, used as a drain follower, provides a very high input impedance required to maintain the very large time constant of the circuit connected to its gate. The stage involving Q9 and Q10 provide additional gain. If R7 is adjusted so that the voltage across R6 is zero, an adjustment of R6 affects only the gain of the stage and not the d-c output level.

(c) Circuit Performance

The circuit of Fig. 3 was used in conjunction with a commercial voltage-controlled oscillator (Wavetek, Model III) . The input carrier frequency was 455 kHz. The voltage levels of the carrier and VCO output were adjusted to be nominally equal and of such magnitude that the PLL bandwidth was 30 kHz. The I. F. filter, centered at the carrier frequency of 455 kHz had an overall bandwidth of 13 kHz. Two counters were used to count individually and separately the positive and the negative spikes. The VCO was initially adjusted to equalize the ratio of occurrence of positive spikes and of negative spikes. The noise level was adjusted to a point where the spike rate, positive and negative together was 200 per minute.

Drift of the VCO was simulated by a slow rotation of the frequency-control dial of the VCO. Figure 4 shows plots of total spike count as a function of the percentage drift of the VCO from its initial setting. In one case the drift-correcting feedback loop was open and in the other case closed. The improvement resulting from feedback is impressive.

II. Modulated Carrier

In the case of the unmodulated carrier we are readily able to use a voltage comparator to mark the occurrence of a spike because the spike amplitude is substantially larger than that of the background smooth noise. When the input is modulated, the spikes are superimposed on the demodulated output signal which may be of comparable amplitude. A further difficulty is to be seen in the following consideration. When, as a result of the modulation, the demodulator output swings, say, in the positive direction, the predominant spike polarity is negative and vice versa. Thus, for example, suppose that the modulation is sinusoidal. Then spikes will occur with greatest likelihood at the peaks of the sinusoidal output signal, since at these peaks the PLL has made the maximum excursions away from the center of range. Further, at the positive peak the spike would be negative and, at the

negative peak, positive. In this sense, the spikes are buried in the signal.

An order of magnitude estimate of the relative amplitudes of signal and spike may be arrived at in the following way. If the baseband bandwidth is f_m and the frequency deviation is Δf , the system bandwidth needs to be $B_{IF} \approx 2(f_m + \Delta f)$. Spikes result from the noise content of this spectral range, and we may therefore expect the spike width to be $\Delta t \sim 1/B_{IF}$. The theory⁽¹⁾ which accounts for the origin of the spikes indicates that on a plot of frequency against time, the spike has an area equal to unity. Hence, the spike has an amplitude which corresponds to a frequency change ΔF at the demodulator input with ΔF determined by $\Delta F \Delta t = 1$. Accordingly

$$\Delta F = 1/\Delta t = 2(f_m + \Delta f)$$

The ratio of the spike amplitude to the peak signal amplitude is

$$\frac{\Delta F}{\Delta f} = 2 \left(\frac{f_m}{\Delta f} + 1 \right) = 2 \left(\frac{\beta + 1}{\beta} \right)$$

in which $\beta = \Delta f/f_m$ is the derivation ratio.

When $\beta \ll 1$, the spikes will stand out sharply against the signal. Such will not be the case when $\beta \gg 1$. For example, if the modulation is sinusoidal and the carrier is modulated through its maximum peak to peak deviation, $2\Delta f$, the spike amplitude will equal the peak to peak signal amplitude. However, if β is large enough, we may again distinguish the spikes in the presence of signal by filtering.

Let the PLL output signal be passed through a high-pass filter of cut-off frequency f_m . The spikes have an energy content which extends through the range $2(f_m + \Delta f) = B_{IF}$. If $B_{IF} \gg f_m$, the signal will be suppressed and the spikes will remain without extensive modification in waveform.

The circuit of Fig.3 with small modification has been used successfully in the case $\beta = 13$. The modulating signal was Gaussian noise, band-limited by a low-pass filter with cut-off frequency at 500 Hz and response down 20 db at 1000Hz. The carrier frequency was again 455 kHz, $B_{IF} = 13\text{kHz}$ and the bandwidth of the PLL was 60kHz. The single essential modification in the circuit was the reduction in magnitude of the coupling capacitor C2. This modification introduces into the circuit a high-pass resistance-capacitance filter network which suppresses the modulation. More elaborate, sharper cutoff filters were also tried with no readily discernible difference in performance. The performance of the circuit with modulation was very nearly identical to the performance shown in Fig. 4 for the unmodulated case.

III Work in Progress

(a) When β is not very large, a high pass filter which serves adequately to suppress the modulation will also differentiate the spike. Therefore, a positive spike will be modified to appear as a positive pulse followed immediately by a negative pulse, while a negative spike will become a negative pulse followed immediately by a positive pulse. In such a case, the system described above fails, because it depends on sensing a difference between the number of negative pulses and positive pulses. It is still possible, however, to distinguish a positive from a negative spike even after differentiation since in one case the differentiated spike makes a positive excursion followed by a negative excursion while for a negative spike the order is reversed. The details of the circuitry needed to reliably detect this difference in proper order is presently being developed.

(b) The feasibility of a procedure for detecting the occurrence of a spike leads to the possibility of developing a scheme for suppressing spikes by cancellation and thereby lowering the threshold. Work is in progress on a scheme for generating at each spike occurrence a pulse opposite in polarity to the spike and of such proportions that the spike will be essentially cancelled. The demodulator output will be transmitted through a delay line to give the spike cancelling circuit adequate time to operate. Preliminary calculations indicated that even a far from complete cancellation will result in a substantial improvement in spike noise. For, if the spikes can be replaced even by doublet type waveform the energy content of the spike noise will be shifted to a frequency range outside the baseband.

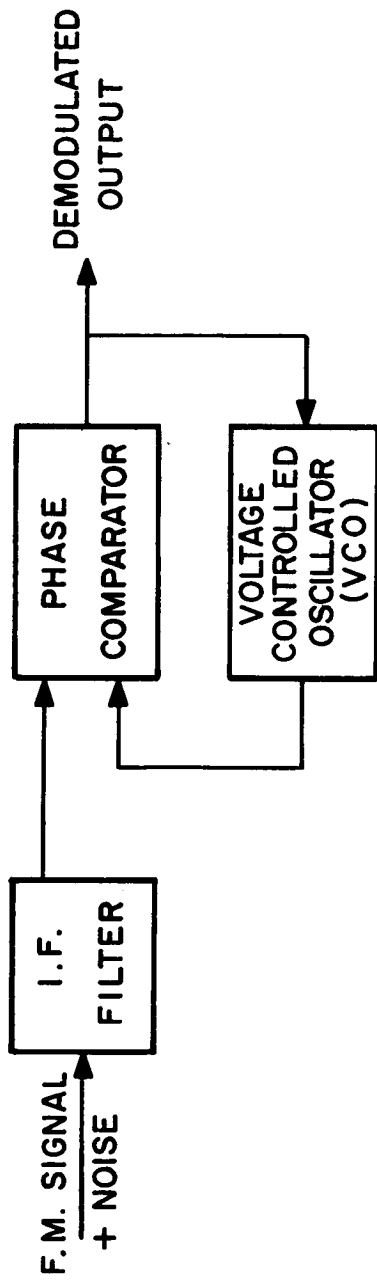


Fig. 1 A phase-locked loop is used to demodulate an F.M. signal

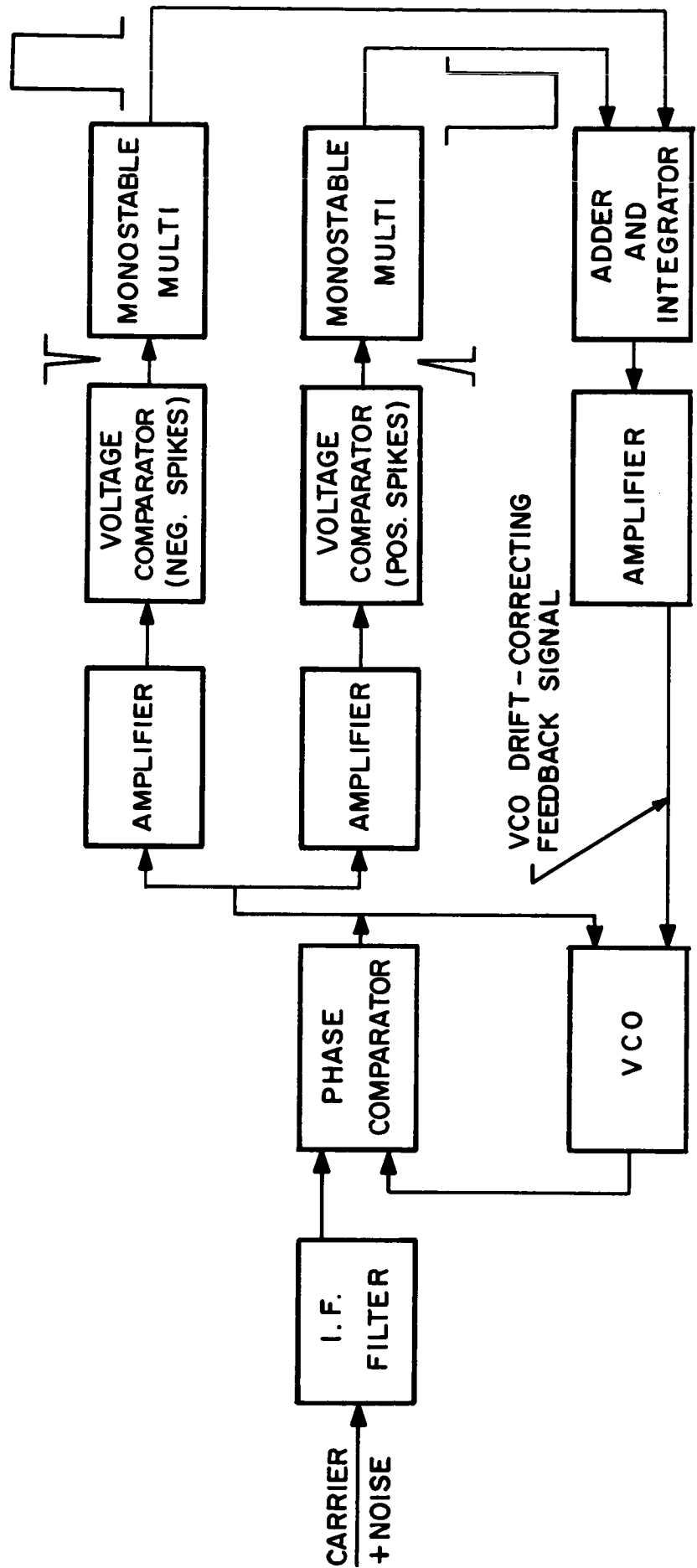


Fig. 2 Block diagram of circuit to correct for VCO drift

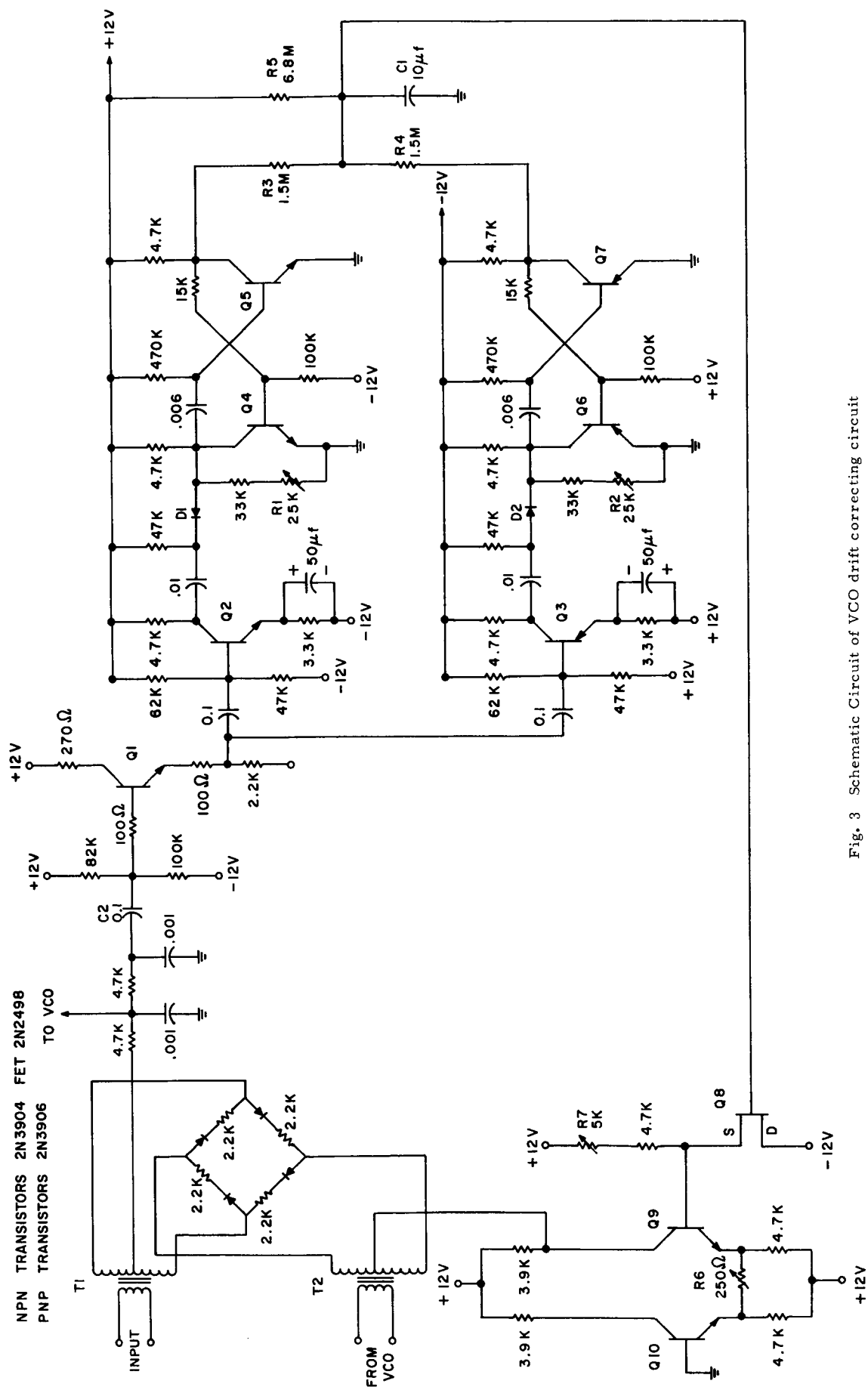


Fig. 3 Schematic Circuit of VCO drift correcting circuit

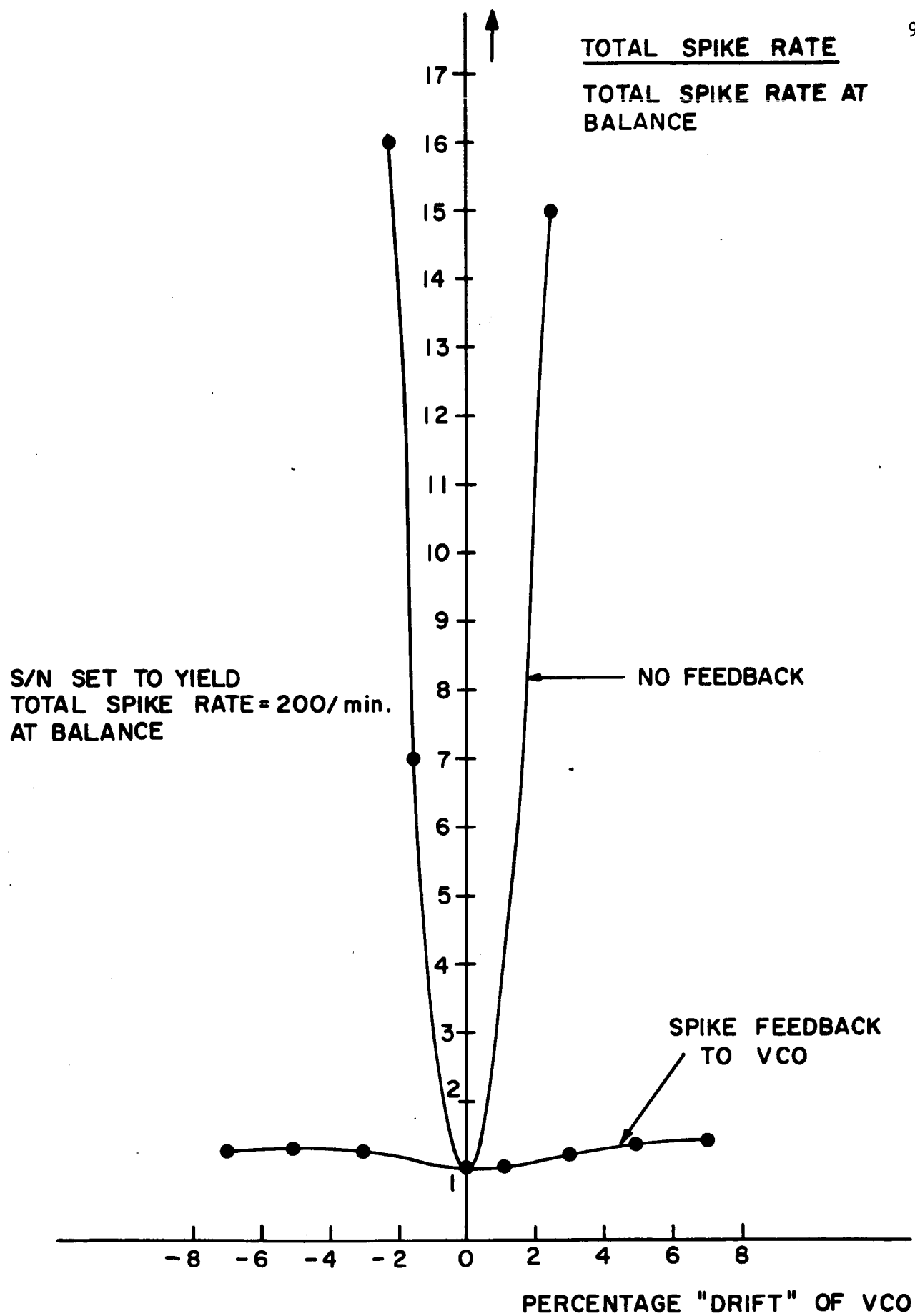


Figure 4

I-C The Frequency Locked Loop

Work is proceeding on the frequency locked loop or FLL upon several fronts simultaneously.

On the one hand the FLL has been shown to provide a limiting case for the older FM with feedback or FMFB circuit. This means that all theoretical and experimental results are directly applicable in providing one starting point in FMFB investigations.

Another aspect of the work has been theoretical. In this work a refined method for calculating both "click" and "gaussian-like" noise outputs from the FLL has been developed. Using this approach computer results have been obtained which allow the calculation of the total noise output from a FLL as well as allowing the individual computation of the several noise components. These components consist of the FLL modified versions of the spike and gaussian-like (parabolic) components normally present in a limiter discriminator output as well as an additional gaussian-like term added by the FLL.

Figure 1 shows these three noise power components for a CNR of 6 db and a ω_{IF}/ω_{MOD} ratio of 2. In addition, it shows the total FLL power and for comparison the total discriminator power for the same conditions. These curves are all for the no modulation case. They are for a loop with an integrator as a low pass filter. Extensive work is under way to extend these results to the modulation case, to consider the case of an ordinary low pass RC network as a loop filter and to compare these results with experimental results. In the experimental case, both direct comparison with a base band FLL fed from a discriminator and comparisons with various types of simulated noise inputs to the FLL are being carried out.

A third aspect of the work has been experimental wherein the loop has been investigated as a threshold extension device for use in receiving both noisy and fading FM signals. Preliminary results have indeed shown that the loop can reduce "spike like" disturbances resulting either from noise or from multipath FM signal interactions

(a) FMFB - FLL Equivalence

If one postulates a FMFB circuit with a wide IF (or a low pass filter that passes the difference frequency from the multiplier and rejects the sum frequency), an FM detector [No limiter] and an integrator for a low pass filter, then it can be shown that this circuit has exactly the same defining equation as the FLL hence the two circuits should yield equivalent performance.

$$e_{in} = a(t) \cos [\omega_1 t + \varphi(t)]$$

$$e_{vco} = B \cos [\omega_2 t + \psi(t)]$$

$$e_{diff} = K_I B a(t) \cos [(\omega_1 - \omega_2) t + (\varphi(t) - \psi(t))]$$

$$e_{disc} = K_{II} a(t) [\dot{\varphi}(t) - \dot{\psi}(t)]$$

$$e_{integrator} = K_{III} \int a(t) [\dot{\varphi}(t) - \dot{\psi}(t)] dt = e_{out}$$

However the VCO is controlled by e_{out} so that

$$\dot{\psi}(t) = K_{IV} e_{out}$$

combining the last two equations

$$\ddot{\psi}(t) + \dot{\psi}(t) K_{III} K_{IV} a(t) = K_{III} K_{IV} a(t) \dot{\varphi}(t)$$

which is exactly the form of the equation for the FLL [Equation 8, Reference 1], hence, this extreme version of a FMFB reduces exactly to a FLL.

[The other extreme case of a FMFB of a narrow IF and a wide low pass filter can be shown to reduce to a phase locked loop] When

$\dot{\psi}(t) K_{III} K_{IV} a(t) \gg \ddot{\psi}(t)$ then $\psi(t) = \varphi(t)$ and distortionless detection occurs.

The postulated FMFB will never suffer loss of lock problems; on the other hand it will not be able to skip input spikes completely but only to reduce them the way a FLL does.

(b) FLL Experimental Results

Figures 2, 3, and 4 show test results for a FLL set-up for the reception of slow scan TV signals via a FM channel. For this test a 100 Hz sinusoidal signal is applied to the $\dot{\varphi}(t)$ input channel while an over modulated { at 1000 Hz } AM carrier is applied to the AM detector channel. Figure 2 shows the detected AM signal as applied to the FET gate. [The DC zero is on the line below the top, the scale is 1v/cm]. By expanding one of the tips, it was seen that the loop hold off time was approximately 220 μ sec/peak. The loop gain (and hence the loop bandwidth) drops to a low value when the com-

bined detected and DC bias signal drops below - 2.3v for this unit.]

Figure 3 shows the results of applying this AM channel signal for the cases where the loop filter has a single pole at 159 Hz. In figure 4 the pole is moved in to (1/11) th of the previous value. In the first case the nominal loop bandwidth is 16kHz which is sufficient to handle the PIB slow scan TV picture when transmitted with a beta of 2.

Note that if a frequency spike had occurred at the instant of the amplitude reduction then the circuit would have produced only a small disturbance instead of a large impulse.

Figure 5 shows the occurrence of a frequency spike in a signal received via the scatterer of Fig. 7 and the PIB water tank channel simulator. It also indicates two regions in which the resultant carrier amplitude became very low. Note that for one small amplitude case a spike occurred [the op going slope near the zero crossing] while for the other [the down going slope near the zero crossing] it did not.

Figure 6 indicates the output of a FLL used to receive this signal. In the lower trace the loop AM channel is disconnected and the click on the rising side of the sine wave is plainly evident. In the upper trace the bias pot setting is reduced from +12v to +9.9v which allows the amplitude minimums to cause holding actions that essentially remove the spike while introducing a slight disturbance upon the opposite slope of the output signal. For most purposes the upper trace would seem to be preferable. In a video signal the effect of this operation would have been to remove either a white or a black dot from a nominal grey area [the color of the interference would depend upon whether up was set up as black or as white].

Reference:

- (1) K.K. Clarke and D.T. Hess "Frequency Locked Loop FM Demodulation" IEEE Transactions on Communications Technology, August 1967, pp. 518-524.

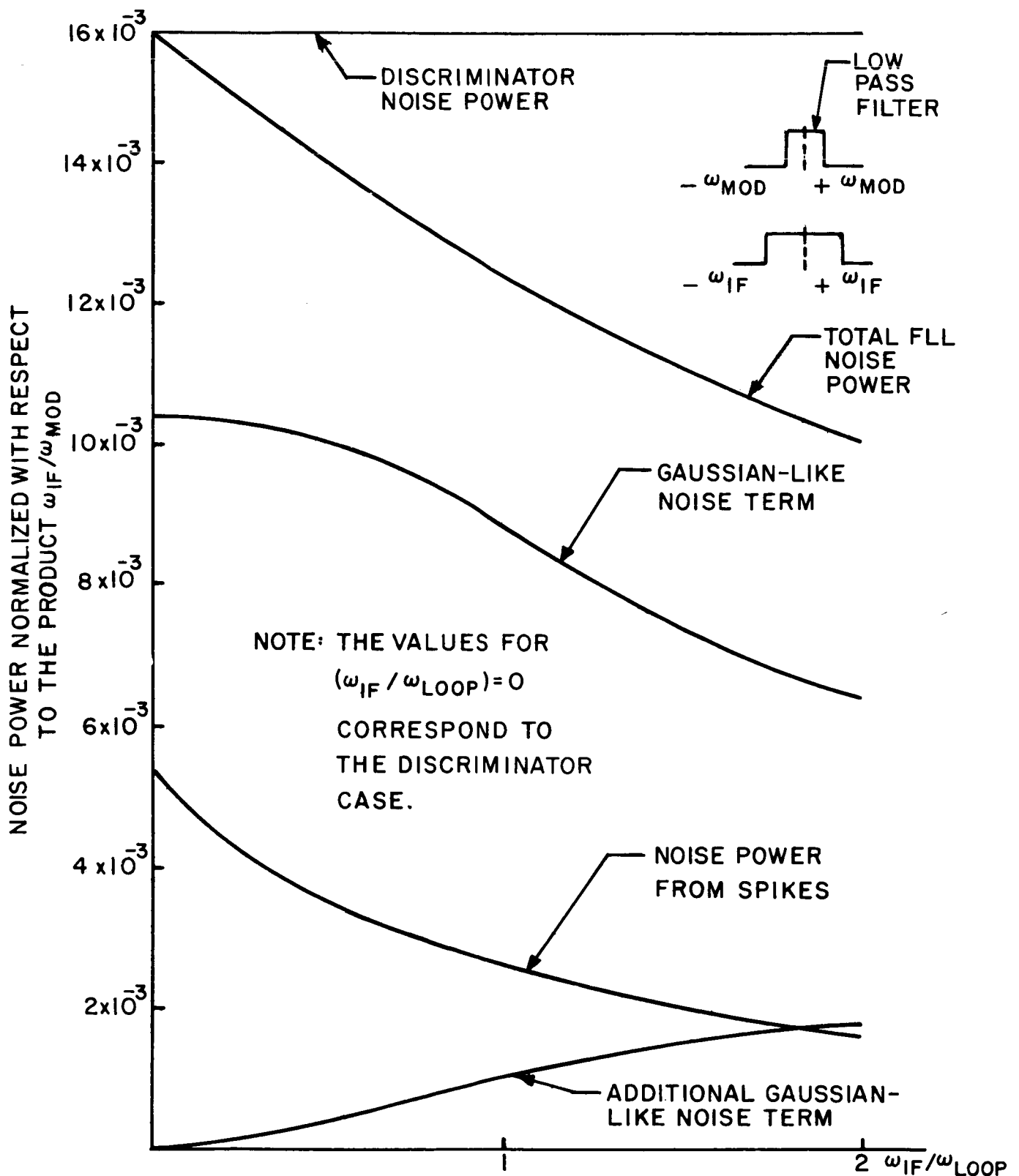


Fig. 1. Frequency locked loop noise power components as a function of $\omega_{IF}/\omega_{Loop}$. $CNR = 6 \text{ dB}$ $[\omega_{IF}/\omega_{Loop}] = 2$. Output low pass filter at ω_{MOD} after the loop. Ideal discriminator output shown for comparison. ω_{IF} and ω_{MOD} as constants. ω_{Loop} as variable

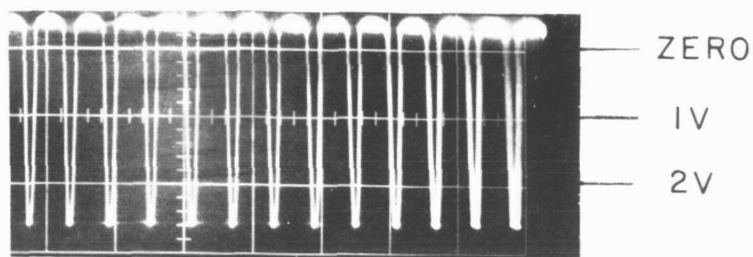


Fig. 2 Simulated amplitude dips as applied to $a(t)$ channel of FLL.

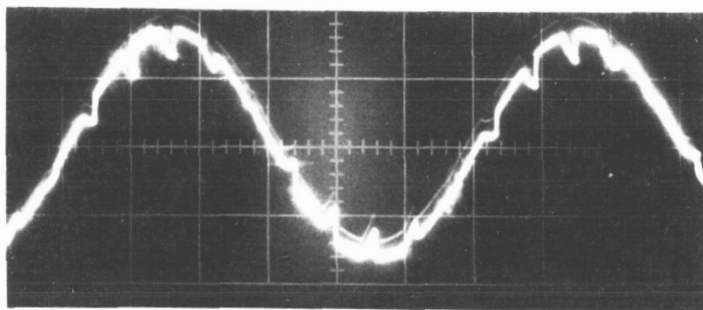


Fig. 3 FLL output for 100 Hz sinusoidal $\phi(t)$ channel input and the $a(t)$ signal of Fig. 2 loop filter at 159 Hz. 1v/cm vertical. Closed loop nominal bandwidth of 15 kHz

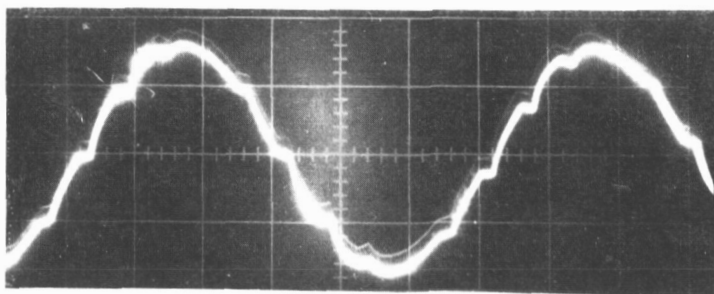


Fig. 4 Same as Fig. 3 except that loop filter has been moved in to 1/11 of previous value. $\omega_{\text{Loop}} \cong 1360 \text{ Hz}$

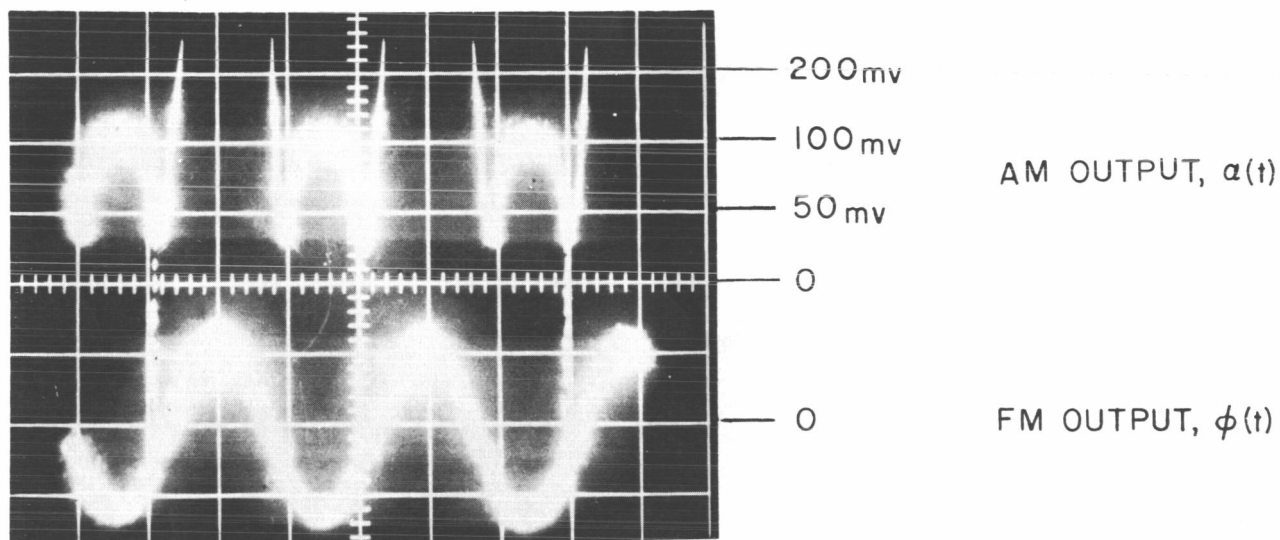


Fig. 5 $a(t)$ and $\phi(t)$ channel inputs for signal transmitted via the PIB water tank simulator. 100 Hz modulating frequency peak deviation of 14 kHz

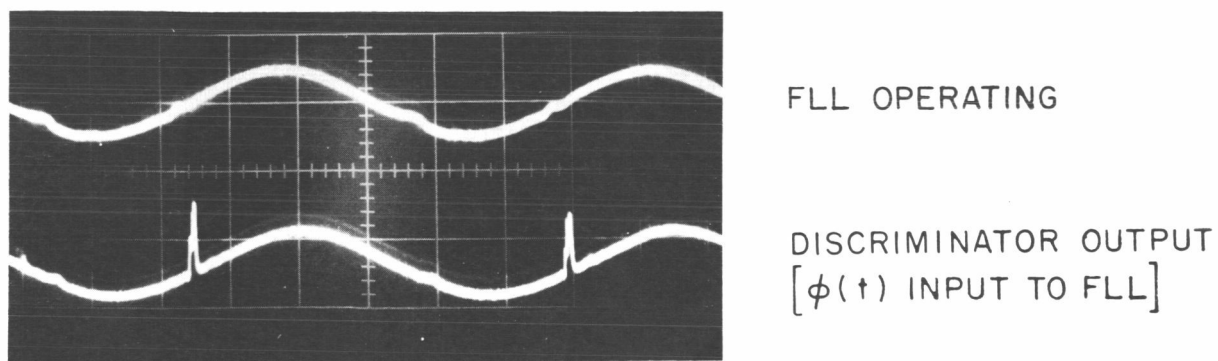


Fig. 6 Effect of FLL in suppressing a "spike" generated by FM transmission in the PIB water tank channel.

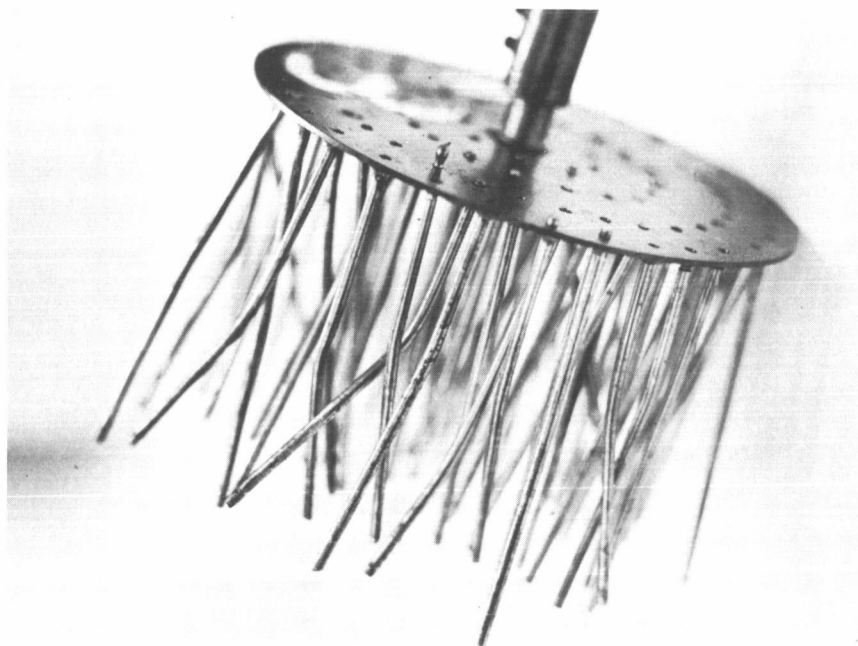


Fig. 7 Scattering element used in the PIB water tank channel simulator to produce the spike of Fig. 6. Glancing operation.

II. PIB Water Tank Channel Simulator

The PIB water tank random channel simulator has been developed to the point where repeatable random channels may be constructed in which envelope probability density, pulse dispersion, and fading rate are independent controllable variables.

As reported earlier^{(1), (2), (3)} the simulator operates at a nominal center frequency of 4 MHz and uses the 2.1×10^5 wave length compression between radio waves in air and sound waves in water to achieve highly directional antennas, small common volumes, and multiple wavelength systems all in a physically small volume.

A schematic representation is shown in Figure 1. Nominal dimensions would include 1 meter between antennas T and R, and water depths of 25 cm. Stray reflections and direct transmissions are kept down 60db from nominal transmission levels through careful grounding procedures and through the use of anti-reflection padding in the tank.

Operation is possible in the glancing mode from antenna T to either or both antennas, R_1 and R_2 , or in a reflecting mode from antenna R_1 via the scattering array to antenna R_2 .

The basic change in operation over the previously reported results in the scattering array shown in Figure 1. Early versions of the channel simulator used air bubbles or mechanical elements (beads) moved by bubbles as scatterers. The current version of the simulator uses arrays of rods or forms as illustrated in Figures 2, 3 and 4.

These arrays are driven by a 20 rmp motor through a gear box with reduction ratios of 1:1, 2:1, 4:1, --- 512:1. Figure 5 shows the motor with one scatterer mounted on it.

Since the probability density of the returned signal depends only upon the array and the path configuration one finds that the fading rate and the probability density have become independent variables. Thus for a given communication system simulation one may control fading rates in steps over a 512/1 range while maintaining the probability density invariant. (The shape of the power spectrum of the fading envelope is not affected by speed changes, the frequency axis is merely expanded or compressed.)

Since the arrays may be removed and replaced (as if they were phonograph records) one may test the same system with different types of fading while maintaining other variables essentially constant.

A very great practical advantage of the speed variation feature is in the measurement of the probability density and of the power spectrum of the fading envelope. These measurements may be made at the highest available speeds so that measurement integration time constants may be relatively short and hence the measurements may be made with commercially available or more simply constructed equipment. (With fading rates of one or two hertz a spectrum analyzer with a bandwidth of 0.01 hertz would be required and measurement times of hours would be necessary to obtain a complete channel average.

Also at slow fading rates digital integration would be necessary for probability density measurements and again hours would be required to obtain one measurement). At the 20 rpm or 10 rpm rates a Quan-Tech Model 304 Wave Analyzer with a 1 h z bandwidth suffices for spectrum measurements while a previously described probability density machine⁽⁴⁾ provides rapid $p(v)$ vs. v measurements.

Average pulse dispersion measurements versus delay are made with a locally designed and constructed circuit of the type shown in Figure 6. This circuit slides a narrow gate along in time and measures the average time spent at each delay. As will be seen in latter illustrations it is possible with this simulator to produce quite non-gaussian dispersion characteristics. In addition to simple average measurements it would be possible to couple two such circuits (with a fixed delay between them) so as to obtain correlation data. It is also desirable to view the returned pulses directly on an oscilloscope so as to note their structure.

Figures 7 and 10 present a set of data of the types just described for the scatterer of Figure 3a. In addition Figure 7 indicates the independence of the outputs of two receiving antennas with centers spaced 2 inches apart (R_1 and R_2 in Figure 1). Thus the simulator is able to provide independently fading signals for use in diversity studies. Figure 10 which shows dispersed pulses received through the channel also indicates different pulse dispersions on the two diversity channels.

Figure 11 indicates the repeatability of the system by indicating the envelope vs. time for more than one complete cycle of revolution of the scatterer of Figure 2 on a glancing path. Fig. 12 indicates the voltage spectrum of the fading envelope for 20 rpm glancing path operation of this scatterer while Fig. 13 indicates the non-gaussian pulse dispersion vs. delay result for in and out operation of the same scatterer.

Work now in progress with the simulator includes:

(a) Investigation of "click" phenomena produced by FM signals transmitted via a random channel. Work includes the identification of such disturbances, the correlation of such disturbances with envelope amplitude variations, the investigation of the distortions introduced by such disturbances in various types of modulation (video, analog, and digital), and the use of threshold extending circuitry such as phase locked loops and frequency locked loops in removing or minimizing such disturbances with envelope amplitude variations, the investigation of the distortions introduced by such disturbances in various types of modulation (video, analog, and digital), and the use of threshold extending circuitry such as phase locked loops and frequency locked loops in removing or minimizing such disturbances.

(b) Investigation of intermodulation and interchannel distortions introduced in multichannel FM signal transmitted via a fading channel. Some theoretical work has been done in this field by others, however, there seems to be no experimental correlation. In addition the theoretical work only treats the case of a delayed channel signal with an amplitude small compared to the main channel. In a severely fading channel this is not a realistic assumption at all.

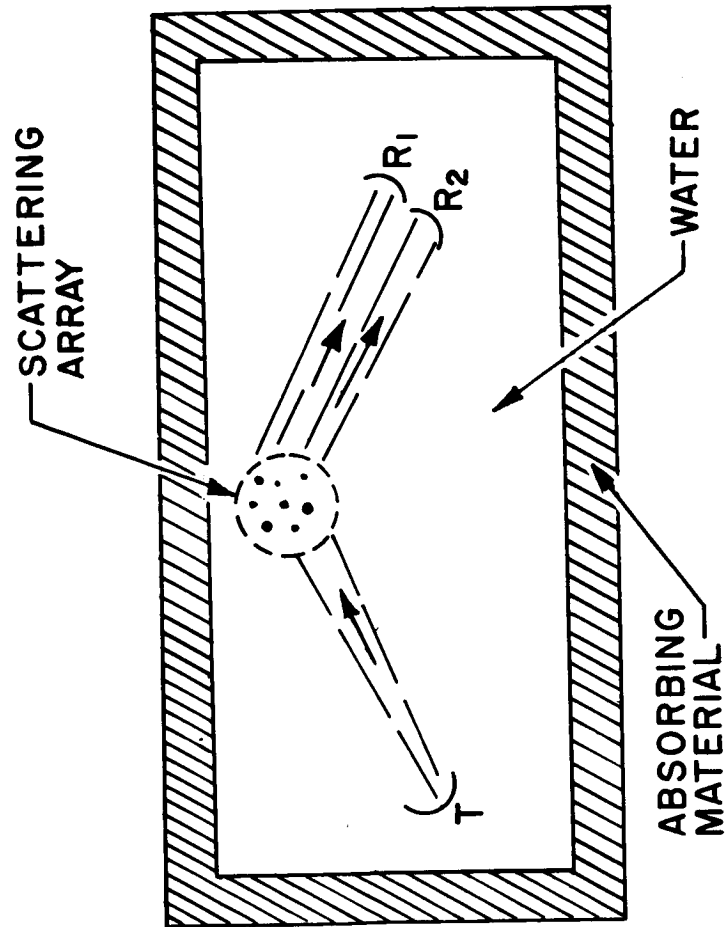


Fig. 1 Schematic top view of water tank channel simulator

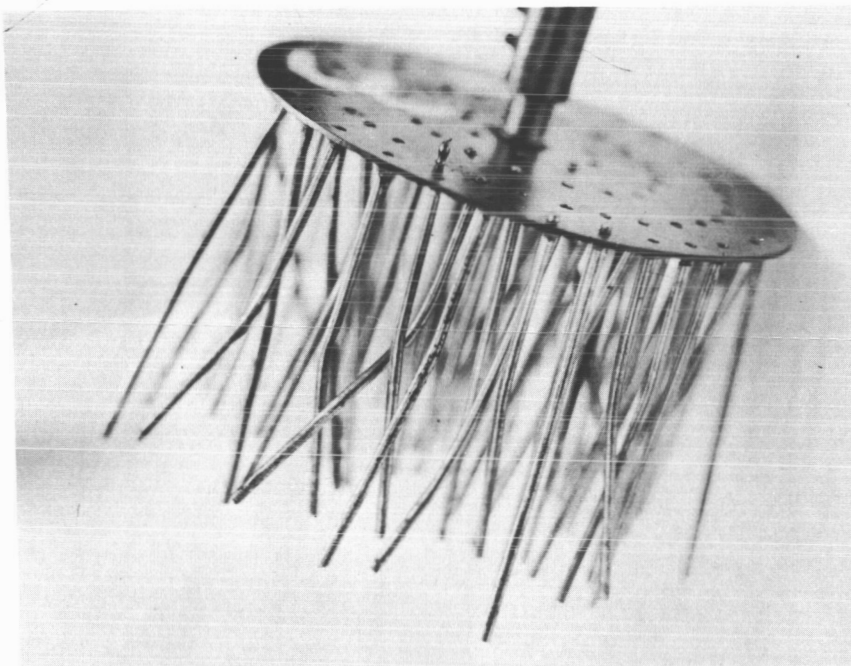


Fig. 2 Multipin Scatterer

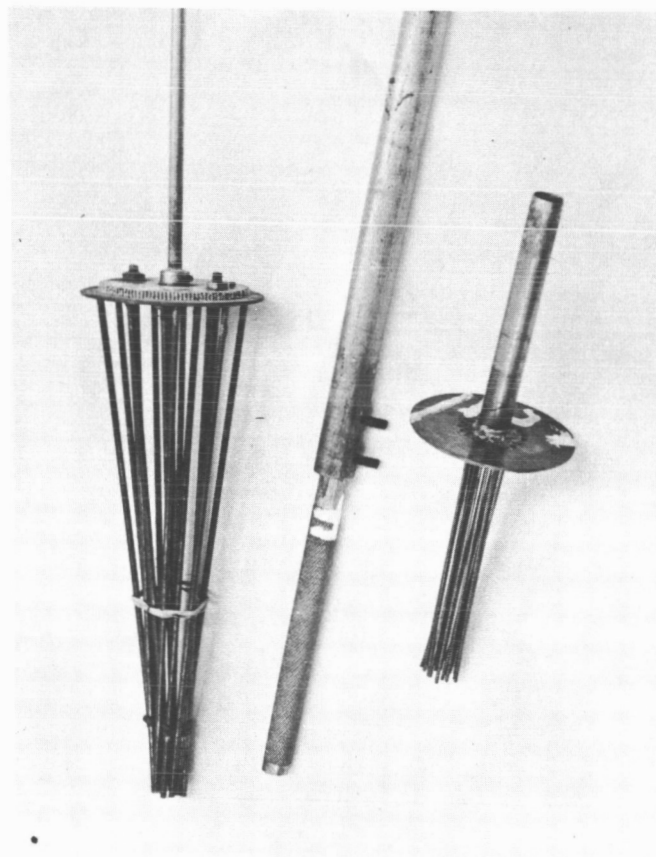


Fig. 3 Three different scattering elements. Labeled a, b, and c from left to right

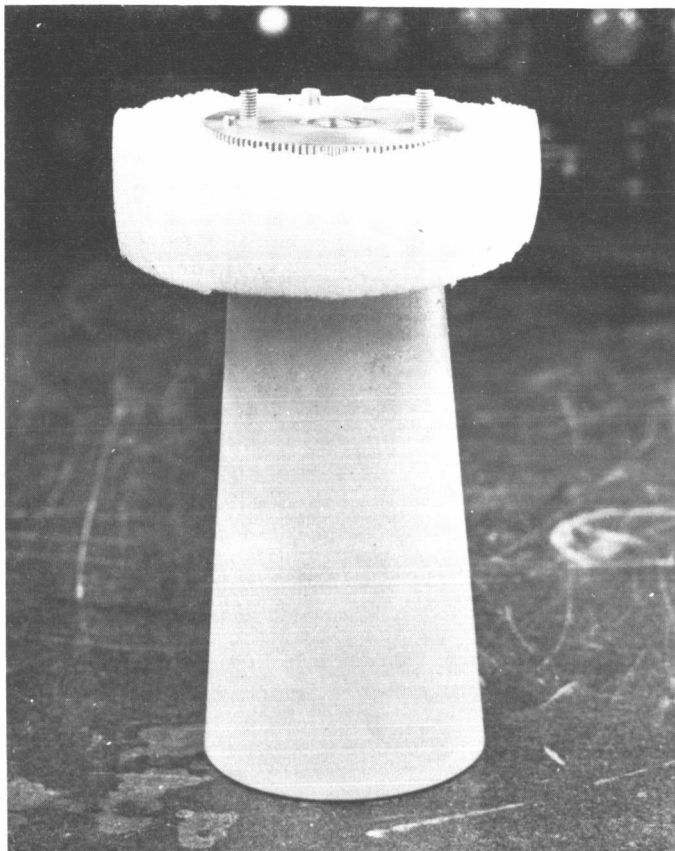


Fig. 4 Rough Plastic Cup
as Scatterer



Fig. 5 Multipin Scatterer with
Variable Speed Motor

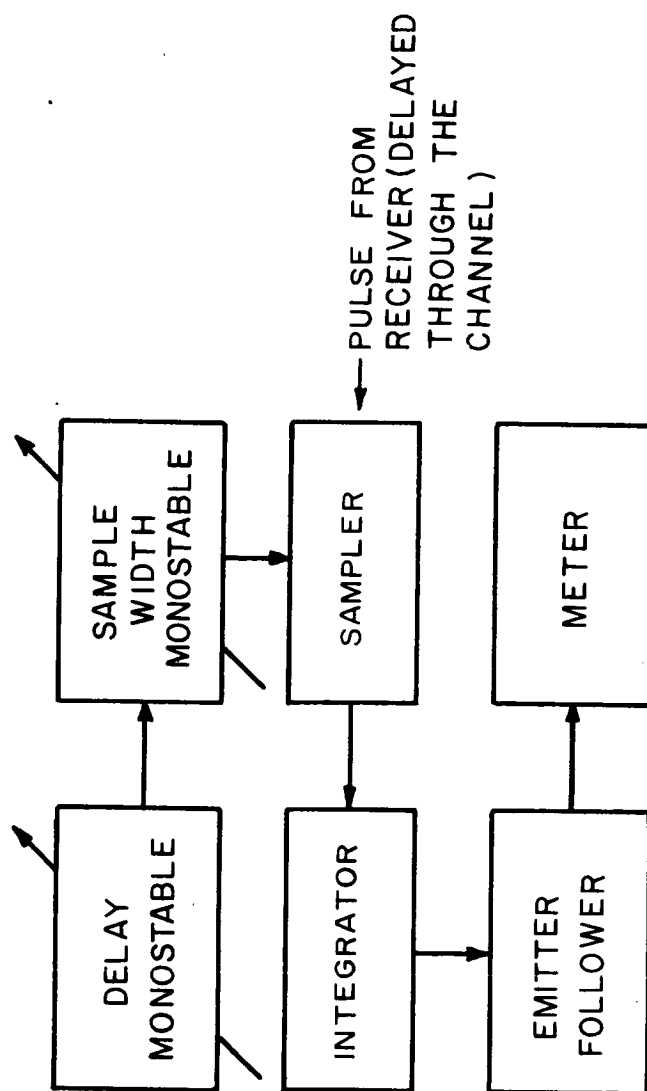


Fig. 6 Average Pulse Dispersion Measurement Circuit

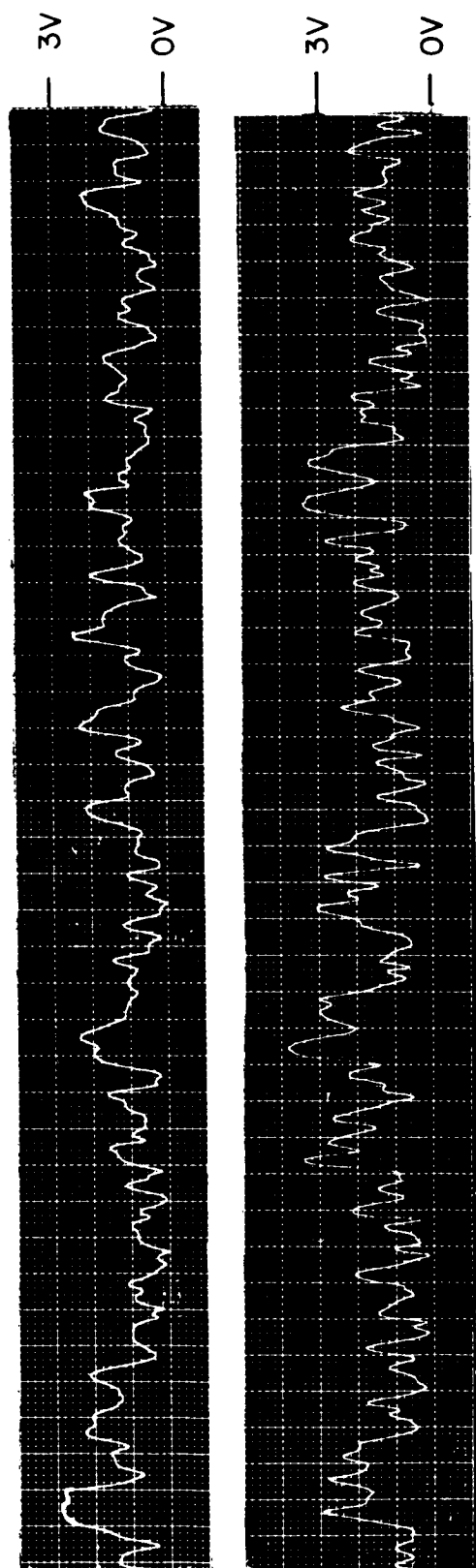


Fig. 7 Chart recordings of the envelopes of simultaneous diversity reception of two 3.9 MHz signals transmitted via array of Fig. 3a 100mm/sec chart speed, 5 RPM rotation rate, Rayleigh Fading

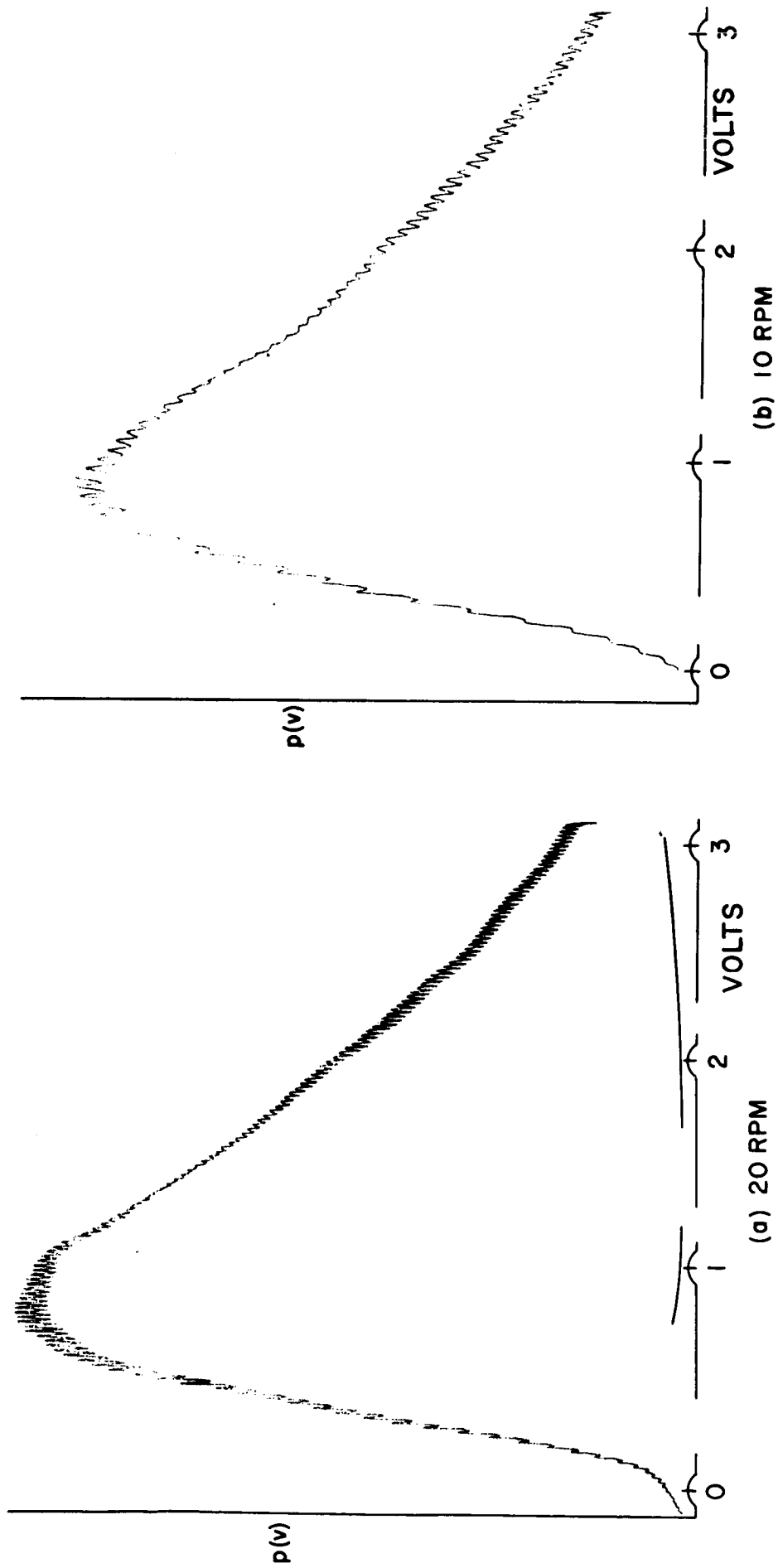


Fig. 8 Probability Densities for Envelopes of Carriers Transmitted via the Scatterer of Fig. 3a

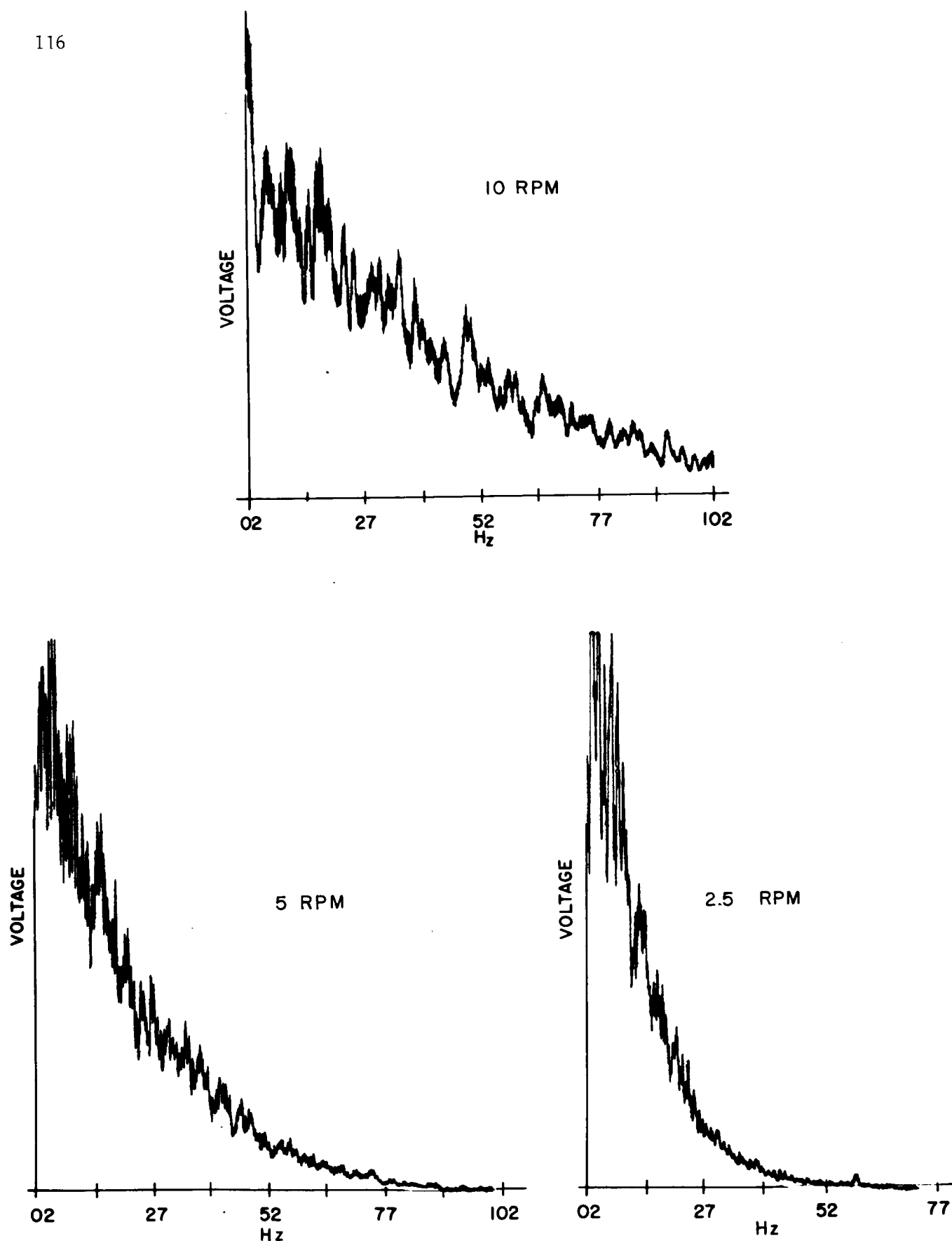
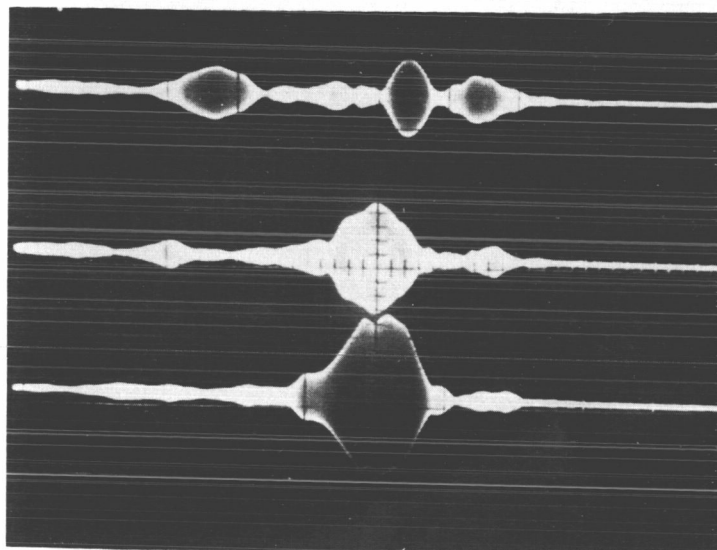
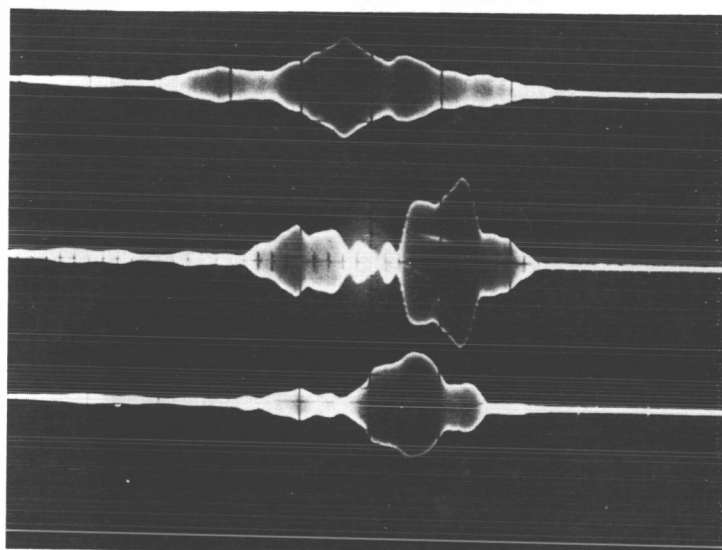


Fig. 9 Voltage Spectra for detected envelopes of 4 MHz carriers transmitted via the scatterer of Fig. 3a. Scatter Rotation rates as indicated. Measured with a Quantech 304 Wave Analyzer.

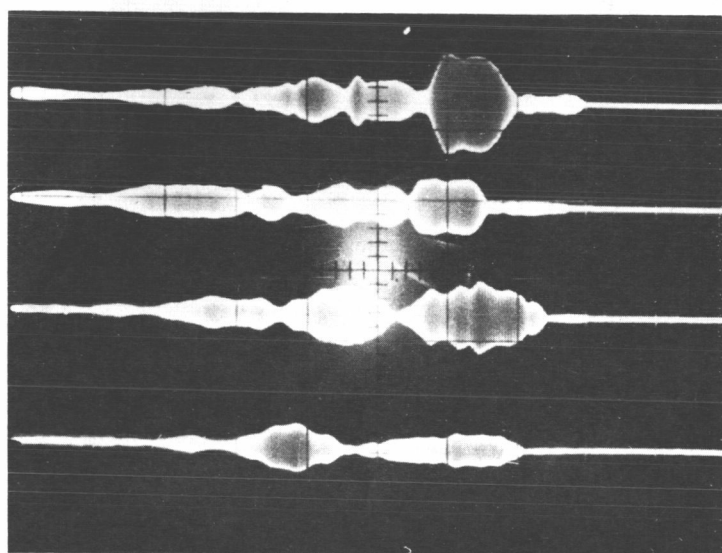


(a) Single channel



(b) Single channel

← TIME



LEFT }
RIGHT }

LEFT }
RIGHT }

(c) Diversity

Fig. 10 Pulse dispersion for $4\mu\text{sec}$ bursts of 4 MHz signal transmitted via the array of figure 3a $4\mu\text{sec/cm}$ horizontally. 0.5v/cm vertically



Fig. 11 Amplitude of fading envelope transmitted via a glancing path and the scatterer of Fig. 2. Chart speed 5mm/sec. Scatterer revolving at (5/8) rpm, 5 volts full scale, zero on bottom line, slightly more than one full revolution shown

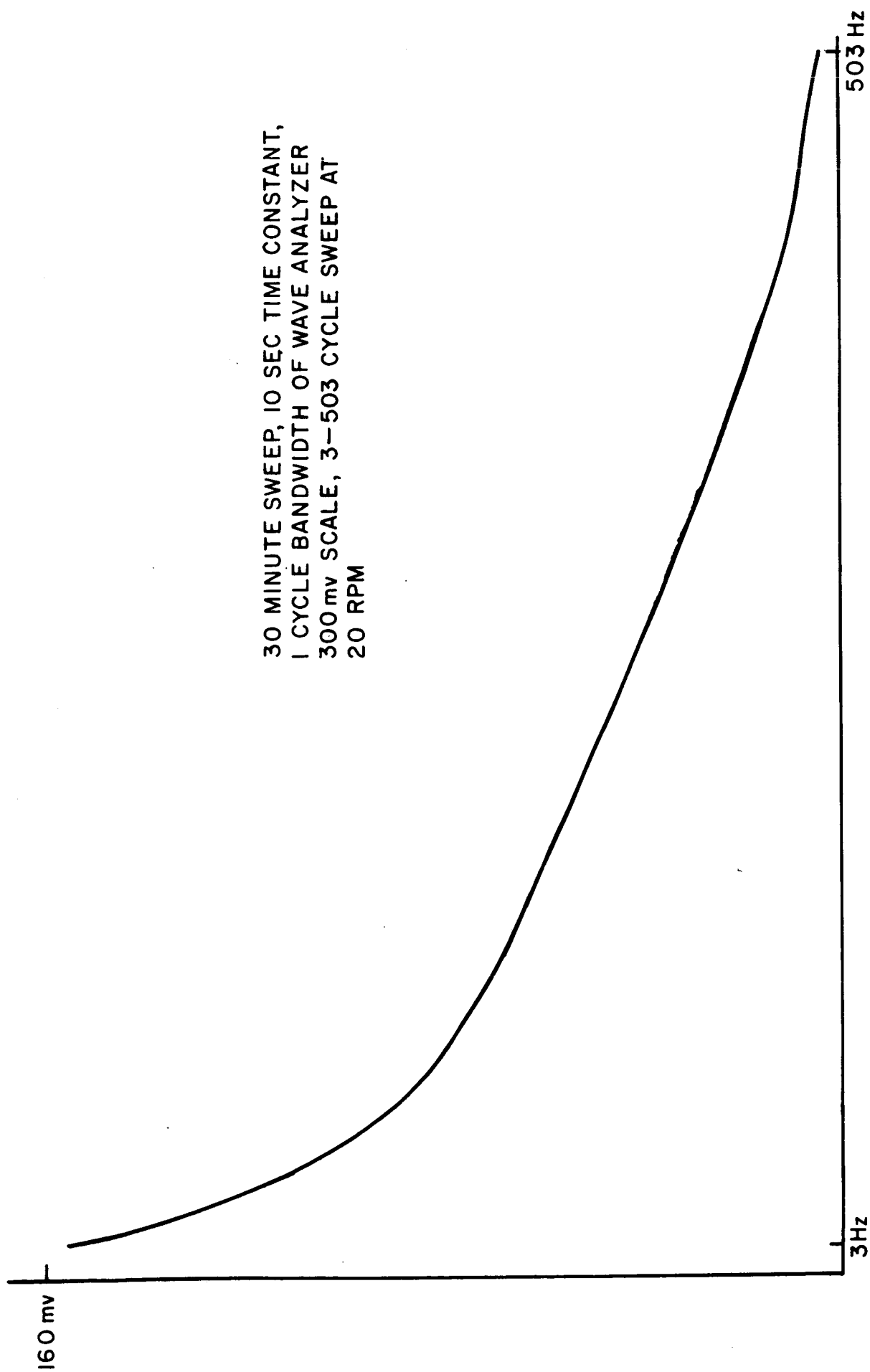


Fig. 12 Voltage Spectrum - Glancing Path-
Scatterer of Fig. 2.

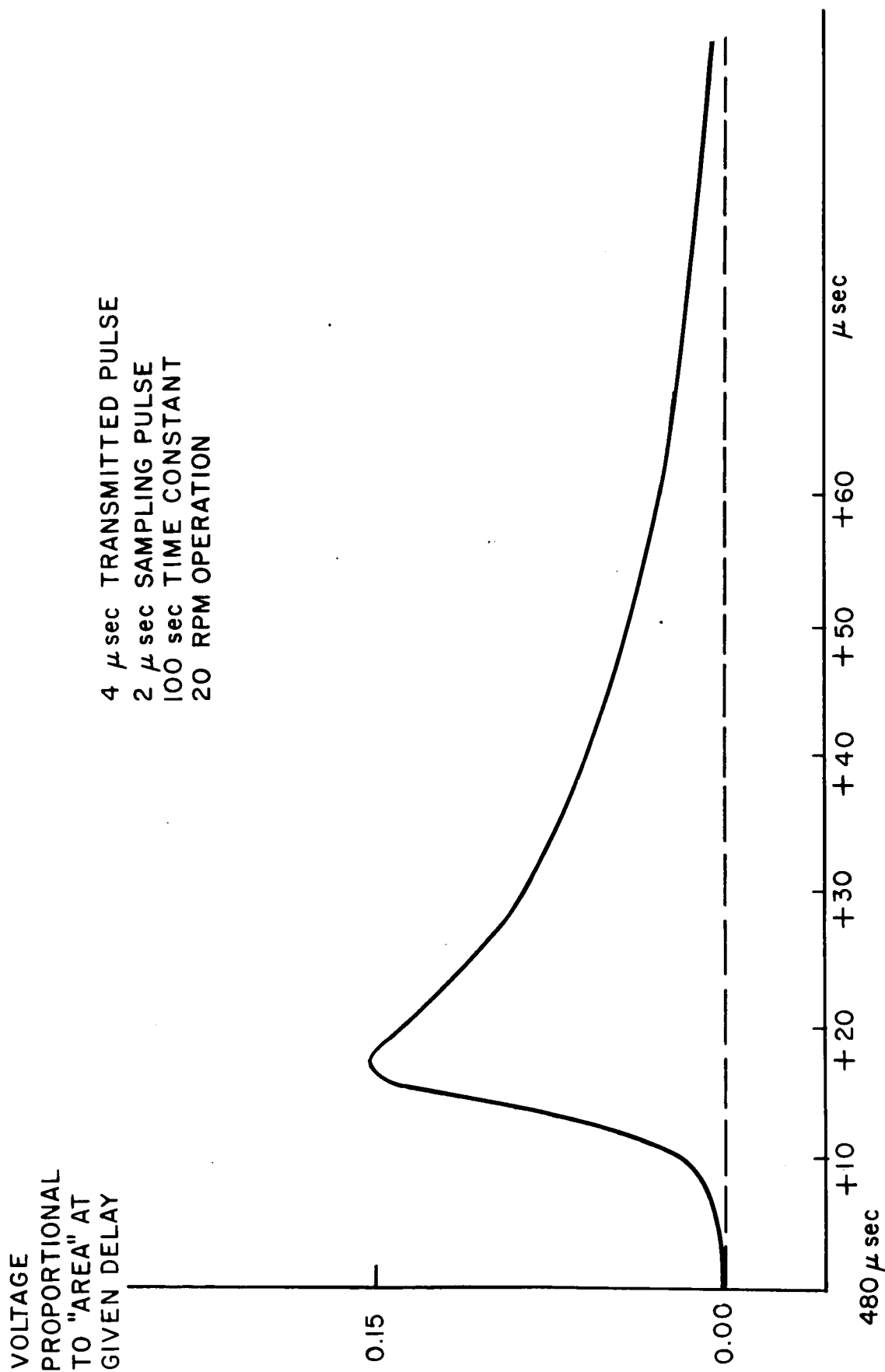


Fig. 13 Average pulse dispersion for channel reflecting from scatterer of Fig. 2. (In and out path)

I - IntroductionA. The Synchronization Problem

The additive white Gaussian noise communication channel is an accurate model of space communication channels [1]. Coherent binary phase shift keyed (PSK) systems operating over such channels yield a 3 db power saving over comparable orthogonal systems such as FSK. The optimum PSK demodulator (Fig. 1) employs a reference signal with perfect phase coherence; such synchronization is unavailable in practical systems due to oscillator phase jitter, doppler, random variation of channel propagation delay, etc. These variations are often slow enough to permit good phase estimation either from the PSK signal itself or from a separately transmitted reference signal. The noisy reference thus obtained, when used in the demodulator of Fig. 1 may result in a probability of error $P(E)$ close to that of the idealized system.

A different but equivalent structure for the correlation detector of Fig. 1 is the product detector shown in Fig. 2.

ω_1, ω_2 are two r.v. defined so that

$$\theta = \tan^{-1} \frac{\omega_2}{\omega_1}$$

This demodulator model more closely parallels the mathematical formulation of the optimum receiver problem (Section II-D).

B. Transmitted Reference and Single Channel Systems

PSK communication systems which operate with a noisy reference signal are called partially coherent systems. They may be classified into two major categories: transmitted reference (TR) and single channel (SC). The transmitted reference systems estimate the phase from a separate reference signal transmitted in addition to the keyed signal. Single channel systems derive their phase estimate directly from the keyed signal. Combination systems are also possible.

The popular TR systems are:

- 1) Adjacent Tone PSK (AT-PSK)
- 2) Quadrature Reference PSK (QR-PSK)

Ideally, these systems deliver identical performance. QR-PSK is

immune to some of the practical problems which beset AT-PSK, but may be more complicated to implement. The major advantage of TR systems is their relative simplicity. Their greatest disadvantage is that a portion of the total available transmitter power must be reserved for transmission of the reference signal, reducing the power available for the keyed signal. Also, the optimum power split between the reference and keyed signals depends on the received signal-to-noise ratio (SNR) which is generally unknown at the transmitter where the adjustment must be made.

The following are SC systems:

- 1) Differential PSK (D-PSK)
- 2) Decision Directed Measurement PSK (DDM-PSK)
- 3) Harmonic Tracking (Squaring) PSK (HT-PSK)
- 4) Costas Synchrolink PSK (CS-PSK).

These systems are not all equivalent. Their common advantage is that the entire transmitter power is used for both data transmission and synchronization. Their major disadvantage is mark-space ambiguity. The absolute phase cannot be established at the receiver; consequently, the derived reference signal can be out of phase by π radians. This necessitates differential data encoding, which generally increases $P(E)$. Encoding methods are discussed in the following section.

C. Absolute and Differential Encoding

Two types of data encoding are employed in binary PSK communication systems; we call them absolute encoding and differential encoding. This terminology does not refer to error correcting coding, but to the way in which the phase shift keyed waveform represents the data sequence. Consider a message consisting of a sequence of plus and minus ones. To form an absolutely encoded PSK signal, we send the phase π to represent a +1, and 0 to represent a -1. To differentially encode the signal, we represent a +1 by a phase change of π radians, and a -1 by no phase change. These two techniques are illustrated in Fig. 3. To illustrate the utility of each scheme, we cite two examples.

First consider a TR system where the receiver tracks a separate sinusoid to obtain a phase estimate. The absolute phase of the keyed signal can be determined at the receiver by comparing the received

reference phase and keyed phase. In this system absolute encoding causes no difficulty. In Section III-C we show that for TR systems absolute encoding always leads to smaller $P(E)$ than does differential encoding.

Now suppose absolute encoding is used in a SC system where a basic ambiguity between 0 and π radians exists. At the beginning of the transmission, a pre-arranged sequence of -1 can be sent to initiate the phase estimation and identify the absolute phase of zero, and there is no decoding problem. However, if at any time during the message transmission the noise causes the phase error of the reference to exceed $\frac{\pi}{2}$ radians, the reference equipment may lose synchronization and relock about π radians. Then the decoded message would be inverted, and $P(E)$ would be near 1. Further, there would be no indication of this condition at the receiver. $P(E)$ of such a system would approach 0.5 for long messages. If differential encoding had been used, this problem would not exist, since the presence or absence of a phase transition between bands can be detected equally well with a reference phase of 0 or π radians. For this reason, differential encoding is almost always used in SC systems. This immunity from reference inversion is bought at the price of increased $P(E)$ (Section III-D), a fact which is ignored in most analyses of SC PSK systems.

II. Transmitted Reference Systems

A. Mathematical Model

The reference measurement time, and hence the quality of the phase estimate, is limited by the rate of change of the received phase, caused by channel variations. This limitation is incorporated into a simple model by postulating a constant received phase, but a limited allowable measurement time. The measurement time qT extends over q bauds of duration T . For convenience in notation, the measurement and data intervals are taken as disjoint, and therefore the noises are independent. In properly designed TR systems, the noises in the two channels are independent even during identical time intervals, so our model is not restrictive. If the signal band is to be included in the phase estimation interval, the same formulas apply with a re-interpretation of q . The received reference and keyed waveforms are:

$$v_r(t) = A \cos(\omega_o t + \theta) + n(t) \quad t \in [-qT, 0]$$

$$v_k(t) = -B \sin(\omega_o t + m_i \frac{\pi}{2} + \theta) + n(t)$$

$$= m_i B \cos(\omega_o t + \theta) + n(t) \quad t \in [0, T]$$

$m_i = \pm 1$ are the (equally likely) data.

θ is a r.v. uniformly distributed on $[-\pi, \pi]$.

$\theta, m_i, n(t)$ are uncorrelated.

$n(t)$ is stationary Gaussian noise with double-sided power spectral

density $S_n(\omega) = \frac{N_o}{2}$.

The total transmitted energy is

$$E = E_r + E_k = \frac{A^2 T}{2} + \frac{B^2 T}{2}$$

The effective SNR's in the reference and data channels are respectively,

$$\alpha = q \frac{E_r}{N_o} \quad \beta = \frac{E_k}{N_o}$$

and the total received SNR is

$$R = \frac{E}{N_0} = \frac{a}{q} + \beta.$$

B. Correlation Detector

The simple correlation detector of Fig. 1 is the optimum receiver for coherent ($\varphi = 0$) binary PSK signals. Now suppose that the reference is noisy, so that φ is a r.v. with non-zero variance. We have succeeded in proving that this detector remains optimum (minimum $P(\epsilon)$), under the following loose conditions on the probability density function (p.d.f.) of the phase error $\varphi = \theta - \hat{\theta}$.

$$1) \quad f(-\varphi) = f(\varphi), \quad \varphi \in [-\pi, \pi]$$

$$2) \quad f(\varphi) \geq f(\pi - \varphi), \quad \varphi \in [0, \pi/2]$$

This result is valid even when $\hat{\theta}$ is not the best possible estimate of θ . We have thus obtained theoretical justification for the intuitive notion that simply replacing the perfect (but unobtainable) phase estimate by whatever estimate is available is a good procedure to follow. The first restriction on $f(\varphi)$, implying that $\hat{\theta}$ is an unbiased estimate of θ , is due to the absence of a natural origin for $f_\theta(\theta)$. Bias in any reasonable estimate is toward the mean value of the a priori p.d.f. of the parameter being estimated. But since θ can be reduced modulo -2π to any 2π interval, our choice of the interval $[-\pi, \pi]$, and consequently of $\bar{\theta} = 0$, is an arbitrary one and should not influence the estimate. The second condition assures that the estimation error will be nearer to 0 radians than to π radians (absence of mark-space ambiguity). To satisfy this condition, for example, it is sufficient (but not necessary) for $f(\varphi)$ to be non-increasing away from zero.

$$f(\varphi_1) \geq f(\varphi_2) \quad \text{for} \quad |\varphi_1| < |\varphi_2|.$$

C. Maximum a' Posteriori Phase Estimator

No one estimation criterion is best for all applications. At this point we have only the indirect (and difficult to apply) criterion that $P(E)$ should be minimum when the estimate is used in the demodulator of Fig. 1. Here we shall discuss the maximum a' posteriori estimator (MAP) and in the next section consider its utility in PSK demodulation.

The MAP computes the phase which was most likely present on the basis of the received reference signal $v_r(t)$ and the a' posteriori p. d. f.'s. Thus $\hat{\theta}$ is the phase which maximizes $f(\theta/v_r)$.

$$f(\theta/v_r) = \frac{f(v_r|\theta)f(\theta)}{f(v_r)} = k' f(v_r|\theta)$$

The phase which maximizes $f(v_r/\theta)$ is known as the maximum likelihood estimate (MLE). Since θ has a uniform a' priori p. d. f., the MAP and MLE are identical for our application. The result for the MAP θ is

$$\hat{\theta} = \tan^{-1} \frac{\omega_2}{\omega_1}$$

where

$$\omega_1 = \int_{-qt}^0 v_r(t) \cos \omega_0 t dt$$

$$\omega_2 = - \int_{-qt}^0 v_r(t) \sin \omega_0 t dt$$

A block diagram of this estimator is shown in Fig. 4.

The p. d. t. of the phase error using this estimate is given by

$$f(\varphi) = \frac{1}{2\pi} e^{-a} + \frac{1}{2} \frac{a}{\pi} e^{-a \sin^2 \varphi} \cos \varphi [\rho + \text{erf}(a \cos \varphi)] \quad \varphi \in [-\pi, \pi]$$

This function is plotted in Fig. 5 for several values of effective SNR a . Since the MAP satisfies both conditions of $f(\varphi)$ given in Section II-B, the detection of Fig. 1 is optimum whenever the phase estimate is of this type.

D. Optimum TR Demodulator

In Section II-B we considered the optimality of the simple detector of Fig. 1 under the assumption that a given phase estimate has already been obtained. Now we seek the demodulator which is optimum in the general case where we have freedom in choosing the criterion for phase estimation. First note that the decision at time $t = T$ is based on the two received waveforms

$$v_r(t) = A \cos(\omega_o t + \theta) + n(t) \quad t \in [-qT, 0]$$

$$v_k(t) = m_i B \cos(\omega_o t + \theta) + n(t) \quad t \in [0, T]$$

To minimize $P(E)$ the maximum a posteriori (MAP) decision rule is used, whereby the most likely message is chosen at each decision instant:

If $f(m_1 | v_k, v_r) > f(m_2 | v_k, v_r)$ decide m_1 ;

otherwise decide m_2 .

Computing the required p.d.f.'s and manipulating the resulting equations, the final decision rule may be written

If $v_1 \omega_1 + v_2 \omega_2 > 0$ decide m_1 where

$$v_1 = \int_0^T v_k(t) \cos \omega_o t dt$$

$$v_2 = - \int_0^T v_k(t) \sin \omega_o t dt$$

$$\omega_1 = \int_{-qt}^0 v_r(t) \cos \omega_o t dt$$

$$\omega_2 = - \int_{-qt}^0 v_r(t) \sin \omega_o t dt .$$

A block diagram of this demodulator is shown in Fig. 6. Comparison with the correlator of Fig. 2 and the phase MAP of Fig. 4 shows that the optimum demodulator for TR PSK is an estimator-correlator where the estimation criterion is MAP.

E. Performance of the Optimum TR System

Several methods have been developed for calculating $P(E)$ for this system. The phase estimate approach, perhaps the most natural one, is considered first.

1. Phase Estimate Approach

First the probability of error is calculated for the correlator of Fig. 1 assuming a constant phase error φ . The result is

$$P(E | \varphi) = \frac{1}{2} \operatorname{erfc}(\sqrt{\beta} \cos \varphi)$$

$$\text{where } \operatorname{erfc} x = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt.$$

When the synchronization is perfect ($\varphi = 0$) we obtain the result for ideal coherent PSK:

$$P_c(E) = \frac{1}{2} \operatorname{erfc} \sqrt{\beta}.$$

For a constant phase error φ , we can define an effective SNR β' as the SNR in a coherent system which would yield the same $P(E)$ as actually obtained.

$$\beta' = \beta \cos^2 \varphi$$

$$\beta'_{\text{db}} = \beta_{\text{db}} - (\sec^2 \varphi)_{\text{db}}$$

The second relationship, which gives the degradation in effective SNR as a function of φ , is shown in Fig. 7. Note that although the degradation is small for small phase offset, at high SNR small degradations can cause large increases in $P(E)$.

In a practical TR PSK system, φ is a random variable, and $P(E)$ can be computed by averaging $P(E | \varphi)$ over the p.d.f. of φ .

$$P(E) = \int P(E|\varphi) f(\varphi) d\varphi.$$

If a suboptimum phase estimate is used, the method remains valid. The optimum estimate is the MAP, with p.d.f. given in Section II-C. Substituting into the integral above and simplifying gives

$$P(E) = \frac{1}{2} \operatorname{erfc} \sqrt{\beta} + \sqrt{\frac{\beta}{\pi}} \int_0^{\pi/2} e^{-\beta \sin^2 \varphi} \cos \varphi \operatorname{erfc} (\sqrt{a} \cos \varphi) d\varphi.$$

Since the first term is $P_c(E)$ for coherent PSK, the second term is the increase in $P(E)$ due to imperfect synchronization. We have also succeeded in obtaining a series expression for $P(E)$ in terms of modified Bessel functions of the first kind $I_n(z)$.

$$P(E) = \frac{1}{2} - \frac{1}{2} \sqrt{a\beta} e^{-\frac{a+\beta}{2}} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \left[I_k\left(\frac{a}{2}\right) + I_{k+1}\left(\frac{a}{2}\right) \right] \left[I_k\left(\frac{\beta}{2}\right) + I_{k+1}\left(\frac{\beta}{2}\right) \right]$$

This expression displays the symmetry of $P(E)$ with respect to a and β which is so strongly suggested by the symmetry of Fig. 6.

2. Unified Analysis of Stein

Stein's unified analysis of binary communication systems [2] can be applied to the TR PSK problem. Briefly, several mathematical transformations change the test on the decision statistic $v_1^{\omega_1} + v_2^{\omega_2}$ to a test between two independent Ricean r.v.'s:

$$P(E) = P(v_+ < v_-)$$

where

$$f_{\pm}(u) = \frac{v}{\sigma^2} \exp \left\{ -\frac{a_{\pm} + v^2}{2\sigma^2} \right\} I_0 \left(\frac{a_{\pm} v}{\sigma^2} \right)$$

$$\sigma^2 = \frac{N_o T}{8}$$

$$a_{\pm} = \frac{|B \pm \sqrt{q} A| T}{2}$$

Stein has computed this probability expression. The result is

$$P(E) = \frac{1}{2} \left[1 - Q \left(\sqrt{\frac{a}{2}} + \sqrt{\frac{\beta}{2}}, \sqrt{\frac{a}{2}} - \sqrt{\frac{\beta}{2}} \right) + Q \left(\sqrt{\frac{a}{2}} - \sqrt{\frac{\beta}{2}}, \sqrt{\frac{a}{2}} + \sqrt{\frac{\beta}{2}} \right) \right]$$

$$= e^{-\frac{a+\beta}{2}} \sum_{m=0}^{\infty} \frac{\epsilon_m}{2} \left(\frac{1-\sqrt{\beta/a}}{1+\sqrt{\beta/a}} \right)^m I_m \left(\frac{a-\beta}{2} \right)$$

where $a \geq \beta$

$$\epsilon_m = \begin{cases} 1 & m = 0 \\ 2 & m \neq 0 \end{cases}$$

and

$$Q(a, b) = \int_b^{\infty} \exp \left\{ -\frac{a^2 + x^2}{2} \right\} I_0(ax) x dx$$

is the Marcum Q-Function.

Another method closely related to this is the so-called characteristic function approach [3]. Here the p.d.f. of the decision statistic is expressed as the Fourier transform of its characteristic function, and this is integrated over the error region. The resulting integrals are complicated and must be evaluated using numerical methods.

F. Design Procedure and Numerical Results

The system design problem for the TR PSK systems studied above consists of determining the optimum power allocation between keyed and reference channels (described by the parameter $\gamma = \frac{Ek}{E}$) and the corresponding minimum $P(E)$, given the measurement parameter q and the SNR, R . No simple formula for γ_{opt} under an average power (R) constraint has been found, so a simple graphical procedure has been devised to estimate γ_{opt} and the associated $P(E)$. Fig. 8 is a plot of β vs a for constant $P(E)$. Its use is illustrated in Fig. 9 for $q = 5$, $R = 8$. The two straight lines

$$\begin{aligned} \beta &= \gamma R \\ \beta &= R - \frac{1}{q} a \end{aligned}$$

describe the relationship of the various system parameters. Their point of intersection is the system "operating point". To determine γ_{opt} , a straight line is drawn from R on the β -axis to qR on the α -axis. The constant $P(E)$ curve which is estimated to be tangent to this line (using logarithmic interpolation between curves) gives the minimum attainable $P(E)$, and γ_{opt} is found by dividing the ordinate of the point of tangency by R . For our example the estimated values are $\gamma_{\text{opt}} = 0.74$, $P(E) = 6.2 \times 10^{-4}$. Note that since normally $q > 1$, the scales of the α - and β -axes are unequal to facilitate the graphical procedure.

Fig. 10 shows the minimum attainable $P(E)$ versus SNR R for several values of q . In Fig. 11 the minimum attainable $P(E)$ is plotted versus q with R as a parameter. The optimum power split $\frac{1-\gamma}{\gamma} = \frac{E_r}{E_k}$ necessary to attain this performance is shown as a function of q with R as a parameter in Fig. 12.

III. Single Channel Systems

A. Mathematical Model

The received waveform is

$$v(t) = m(t) C \cos (\omega_0 t + \theta) + n(t) \quad t \in [-qT, T]$$

where $m(t)$ is a square wave taking either of the values ± 1 with equal probability and independently in each band, and θ and $n(t)$ were specified in Section II-A. The SNR is

$$R = \frac{E}{N_0} = \frac{C^2 T}{2 N_0},$$

and differential encoding is used for reasons discussed in Section I-C. The structure of the optimum receiver for an SC system depends upon the amount and complexity of signal processing that can be done in the receiver. Studies of SC systems are usually concerned with the performance of near optimum but easily implemented receivers, rather than with very complicated optimum receivers.

B. Differential PSK

Differential PSK is the simplest SC system, and corresponds to DDM-PSK with $q = 1$. Since the decision is based on a comparison of the signals received in the present and previous band, no decision feedback is actually necessary.

1. Average Probability of Error

$P(E)$ can be calculated using the phase estimate approach of Section II-E. Since the present baud constitutes the keyed signal, and the previous band the reference signal, we have for D-PSK

$$\alpha = R$$

$$\beta = R$$

$$P(E | \varphi) = \frac{1}{2} \operatorname{erfc} (\sqrt{R} \cos \varphi)$$

$$f(\varphi) = \frac{1}{2\pi} e^{-R} + \frac{1}{2} \sqrt{\frac{R}{\pi}} e^{-R \sin^2 \varphi} \cos \varphi [1 + \operatorname{erf}(\sqrt{R} \cos \varphi)] \quad \varphi \in [-\pi, \pi]$$

$$P(E) = \int P(E | \varphi) f(\varphi) d\varphi$$

This integral can be solved directly, with the help of the fact

$$\int f(\varphi) d\varphi = 1,$$

to yield the well-known result

$$P(E) = \frac{1}{2} e^{-R}.$$

2. Probability of Double Errors

The fact that each signal baud is involved in two decisions in D-PSK leads to the speculation that errors may tend to occur in pairs. The principal quantity of interest is the probability that an error will occur in baud $j+1$ given that an error has occurred in baud j .

$$P(E_{j+1} | E_j) = \frac{P(E_{j+1}, E_j)}{P(E_j)} = 2 e^R P(E_{j+1}, E_j).$$

Salz and Saltzberg [4] obtained the following result:

$$P(E_{j+1}, E_j) = \frac{1}{4\pi} e^{-R} \int_0^\pi [1 - \operatorname{erf}(\sqrt{R} \cos \theta)]^2 \{1 + \sqrt{R} \pi \cos \theta e^{R \cos^2 \theta} [1 + \operatorname{erf}(\sqrt{R} \cos \theta)]\} d\theta$$

Their method of derivation was too complicated to give any understanding of the processes involved. First they derived the joint p.d.f. of two successive decision statistics in the form of a double integral. Then this p.d.f. was integrated over the quadrant corresponding to two errors. The resulting four-fold integral was finally manipulated into the above form. Below we rederive this result in a simpler manner which makes direct use of the system characteristics which cause the double errors.

The method used for double errors is related to the phase estimate approach used above to calculate $P(E)$. Since each decision is based on the comparison of two adjacent bauds, the roles of reference baud and data baud are not uniquely determined because the information is contained in the presence or absence of a phase transition between the bauds. Recall the symmetry of $P(E)$ with respect to α and β demonstrated in Section II-E, and note the obvious symmetry of the optimum receiver ($q = 1$ for D-PSK) of Fig. 6. Fig. 13a shows three signal bands and the two decisions D_1 , D_2 made on the basis of presence or absence of transitions at T_1 , T_2 . The correlation between these decisions is due to the data for D_1 being identical to the reference for D_2 . Now consider interchanging the labels of reference and data for D_1 (Fig. 13b). The same decision will be reached because of its symmetrical dependence on reference and data (T_1 is not altered). The simplification thus obtained is that both decisions use the same (nominal) reference. Since the data bauds are distinct, their respective probabilities of error conditioned on φ (derived from baud 2) are independent. Then we have simply

$$\begin{aligned} P(E_{j+1}, E_j) &= \int P(E_{j+1} | \varphi) P(E_j | \varphi) f(\varphi) d\varphi \\ &= \int \left[\frac{1}{2} \operatorname{erfc}(\sqrt{R} \cos \varphi) \right]^2 f(\varphi) d\varphi \end{aligned}$$

with $f(\varphi)$ as given previously. This result is identical to that of Salz and Saltzberg, is more simply derived, and shows that the frequency of double errors in D-PSK can be directly attributed to the shared band of adjacent decisions.

C. Degradation Due to Differential Encoding

To facilitate our comparison of TR and SC systems in the next section, we examine here the effect of differential encoding on TR system operation. The adjacent-bit dependence of differential encoding makes it necessary to identify four states of a TR system. In Fig. 14 we show the four states, corresponding to present and past detections.

For absolute encoding, correct detection implies correct decoding. When differential encoding is used, however, the decoding depends on the previous detection as well. The relationship between detection and decoding is shown in the table accompanying the state diagram. The detection probabilities are

$$q_1 = P(E_{j+1} | C_j)$$

$$q_2 = P(E_{j+1} | E_j)$$

$$p_i = 1 - q_i \quad i = 1, 2.$$

The probability state vector and probability transition matrix are (from Fig. 14)

$$\underline{P}_j = \begin{bmatrix} P_{1j} \\ P_{2j} \\ P_{3j} \\ P_{4j} \end{bmatrix} \quad \underline{z} = \begin{bmatrix} p_1 & p_1 & 0 & 0 \\ 0 & 0 & p_2 & p_2 \\ q_1 & q_1 & 0 & 0 \\ 0 & 0 & q_2 & q_2 \end{bmatrix}$$

If at the j^{th} decision the probability of the states is \underline{P}_j , the state probabilities at the $j+1$ st decision are

$$\underline{P}_{j+1} = \underline{z} \underline{P}_j$$

Since the state diagram is aperiodic and irreducible, a steady-state probability state vector exists [5] and satisfies the equation

$$\underline{P}_\infty = \underline{z} \underline{P}_{cs} \quad \underline{P}_{cs} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix}$$

We obtain the set of equations

$$P_1 = p_1 (P_1 + P_2)$$

$$P_2 = p_2 (P_3 + P_4)$$

$$P_3 = q_1 (P_1 + P_2)$$

$$P_4 = q_2 (P_3 + P_4)$$

The solution of these equations is

$$P_1 = \frac{p_1 p_2}{q_1 + p_2}$$

$$P_2 = \frac{q_1 p_2}{q_1 + p_2}$$

$$P_3 = \frac{q_1 p_2}{q_1 + p_2}$$

$$P_4 = \frac{q_1 q_2}{q_1 + p_2}$$

For absolute and differential encoding, the average probabilities of error are, respectively,

$$P_a(E) = P_3 + P_4 = \frac{q_1}{q_1 + p_2}$$

$$P_d(E) = P_2 + P_3 = \frac{2q_1 p_2}{q_1 + p_2}$$

Now note that in a reasonable system, the conditional detection probability

$$p_2 = P(C_{j+1} | E_j) > \frac{1}{2}$$

and

$$P_d(E) > P_a(E)$$

Thus differential encoding always yields a larger $P(E)$ than does absolute encoding in a TR system. Also, in a system where

adjacent detection probabilities are independent, the results simplify to

$$q_1 = q_2 = q$$

$$P_a(E) = q$$

$$P_d(E) = 2q(1-q).$$

D. Comparison of TR and SC Systems

For simplicity we shall focus attention on two specific cases. These are sufficient to demonstrate that previous analyses of SC systems resulted in incorrect conclusions because the necessity of differential encoding was ignored. We show that for low values of q , SC systems are superior, while for very large q , TR systems perform better.

1. $q = 1$

The SC system which we consider for $q = 1$ is D-PSK. We have the well-known result

$$P_S(E) = \frac{1}{2} e^{-R}$$

A comparable TR system has either $q = 1$ or $q = 2$. The second case arises when the reference is derived over the present as well as previous reference band. SC systems cannot increase the reference SNR by using the present data band. For $q = 1$ we have the exact result

$$P_{T1}(E) = \frac{1}{2} e^{-\frac{R}{2}}$$

while for $q = 2$ computer evaluation and optimization of a $P(E)$ expression from Section II-E is necessary. These three probabilities are shown in Fig. 15 as a function of SNR, R . The SC system is superior to both TR systems.

2. $q = \infty$

As the measurement time increases without bound, a perfect reference is obtained with either SC or TR systems. For the TR system a negligible amount of power is sufficient for reference transmission. The performance of the absolutely encoded TR system with perfect $q = \infty$ is, from Section II-E,

$$P_{T\infty} = \frac{1}{2} \operatorname{erfc} \sqrt{R}$$

For the SC system, the perfect reference signal implies that adjacent detection decisions are independent, so it is equivalent to a differentially encoded TR system. The result of Section III-C can therefore be used.

$$P_{S\infty} = \operatorname{erfc} \sqrt{R} \left[1 - \frac{1}{2} \operatorname{erfc} \sqrt{R} \right].$$

These results are plotted in Fig. 16. For very large q , we see that $P(E)$ of SC systems is nearly double that of TR systems.

The above comparison of SC and TR systems shows that the need for differential encoding in SC systems degrades the $P(E)$ appreciably, and for large measurement times, TR systems using absolute encoding are superior to SC systems. This is in contrast with a result of Van Trees [6], who claimed that SC systems are always superior. He neglected the differential encoding in his analysis.

References

1. Viterbi, A. J., Principles of Coherent Communication, McGraw-Hill Book Company, 1966, p. 5.
2. Stein, S., "Unified analysis of certain coherent and noncoherent binary communications systems," IEEE Trans. on Information Theory, January 1964, pp. 43-51.
3. Bussgang, J. J., and M. Leiter, "Error rate approximations for differential phase-shift keying," IEEE Trans. on Communications Systems, March 1964, pp. 18-27.
4. Salz, J., and B. R. Saltzberg, "Double error rates in differentially coherent phase systems," IEEE Trans. on Communications Systems, June 1964, pp. 202-205.
5. Fano, R. M., Transmission of Information, The M. I. T. Press, 1961, p. 107.
6. Van Trees, H. L., "Optimum power division in coherent communication systems," IEEE Trans. on Space Electronics and Telemetry, March 1964, pp. 1-9.

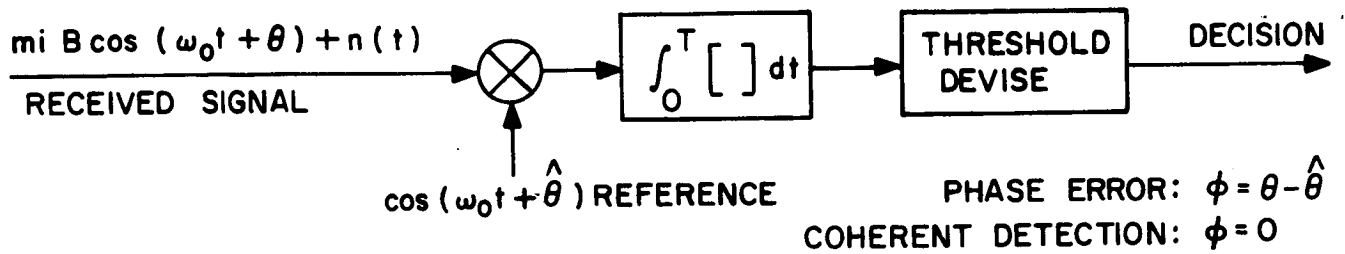


Fig. 1 Correlation detector

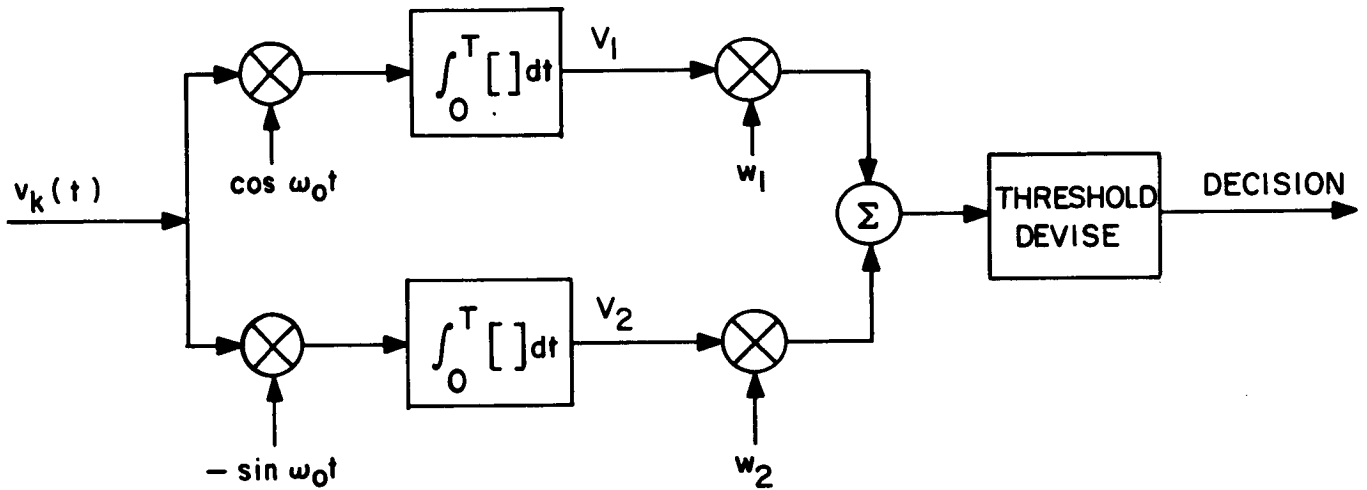


Fig. 2 Product detector

MESSAGE SEQUENCE		+1	+1	-1	+1	-1	-1	-1
ABSOLUTE ENCODING		π	π	0	π	0	0	0
DIFFERENTIAL ENCODING	0	π	0	0	π	π	π	π

Fig. 3 Absolute and differential encoding

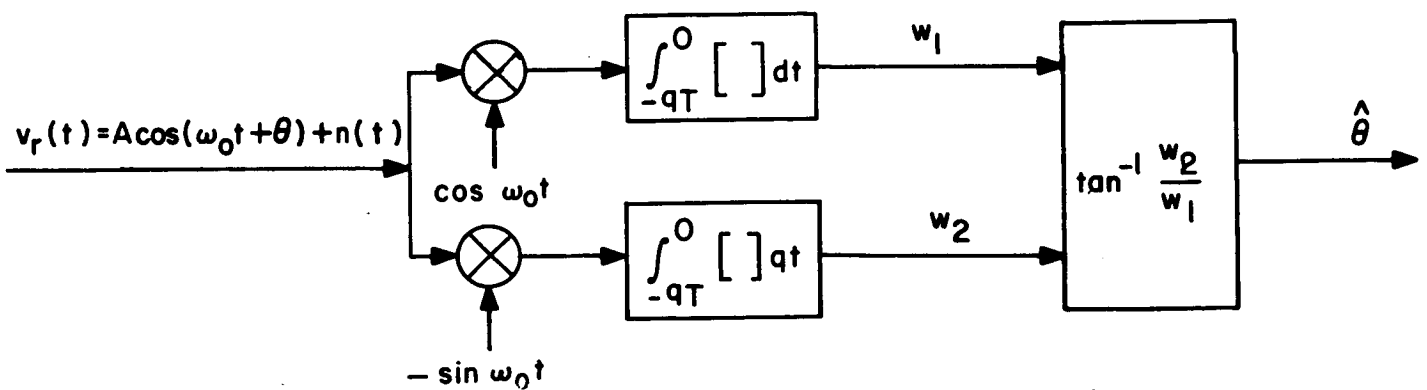


Fig. 4 MAP phase estimator

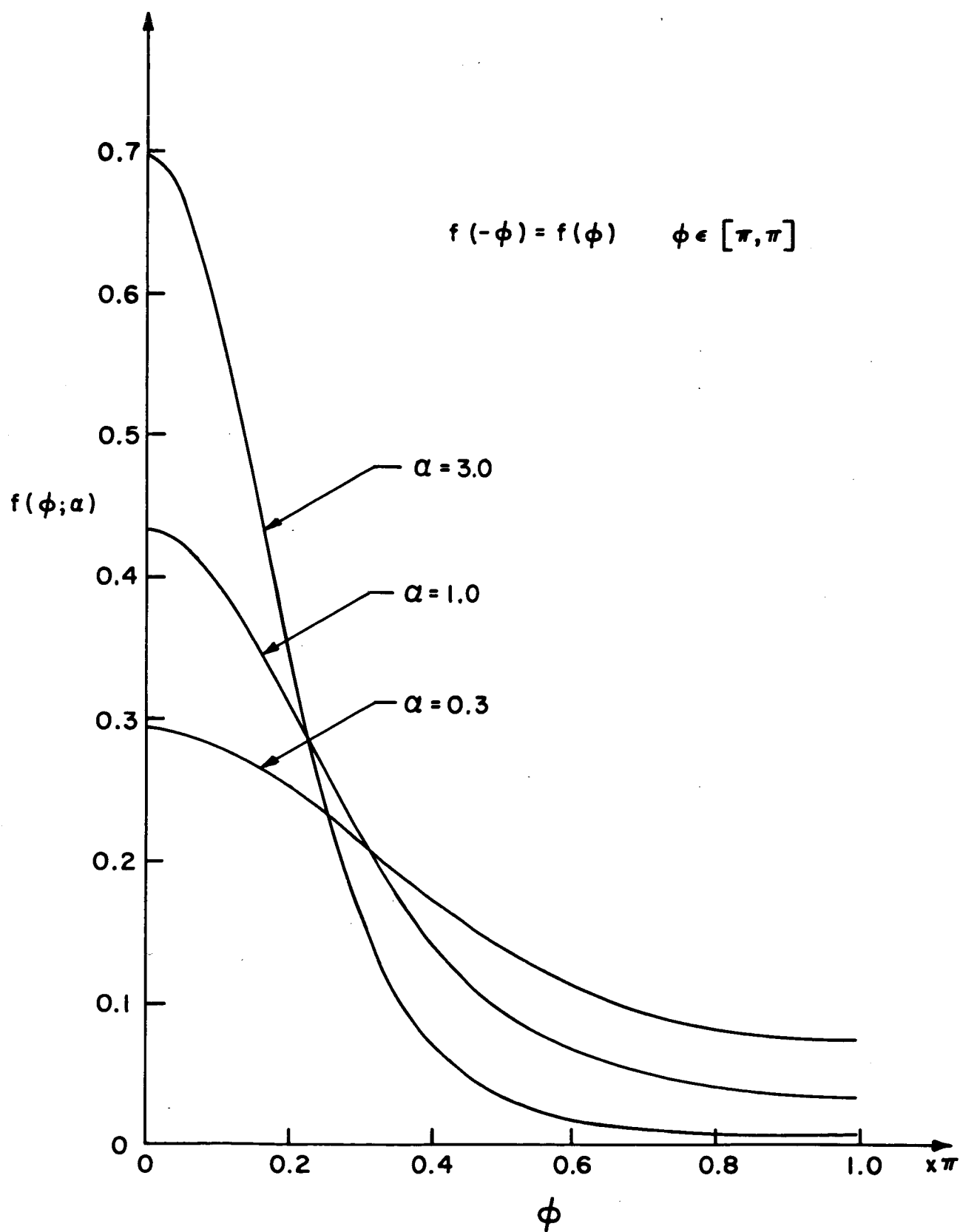


Fig. 5

P.D.F. of phase error of MAPE

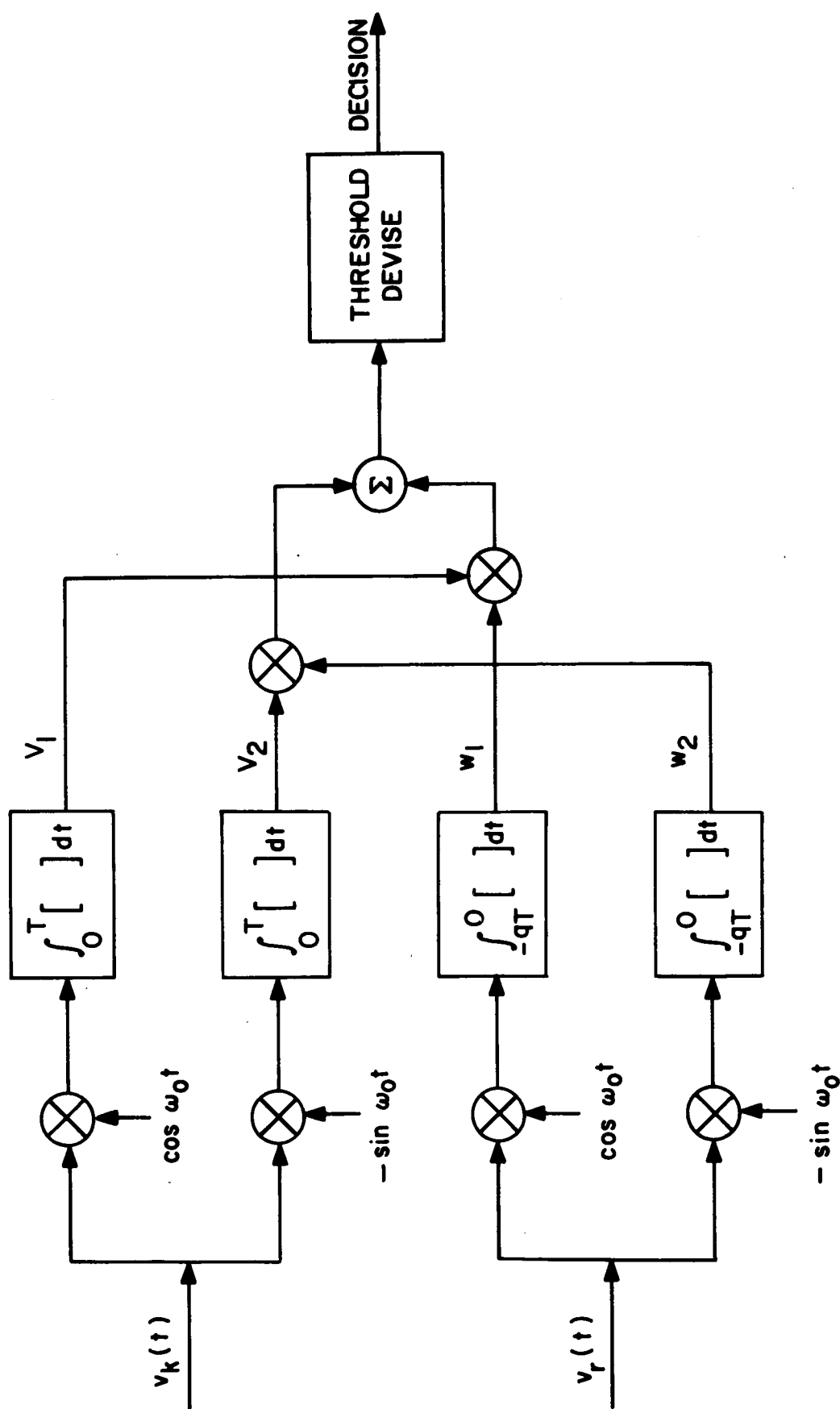


Fig. 6 Optimum TR PSK demodulator

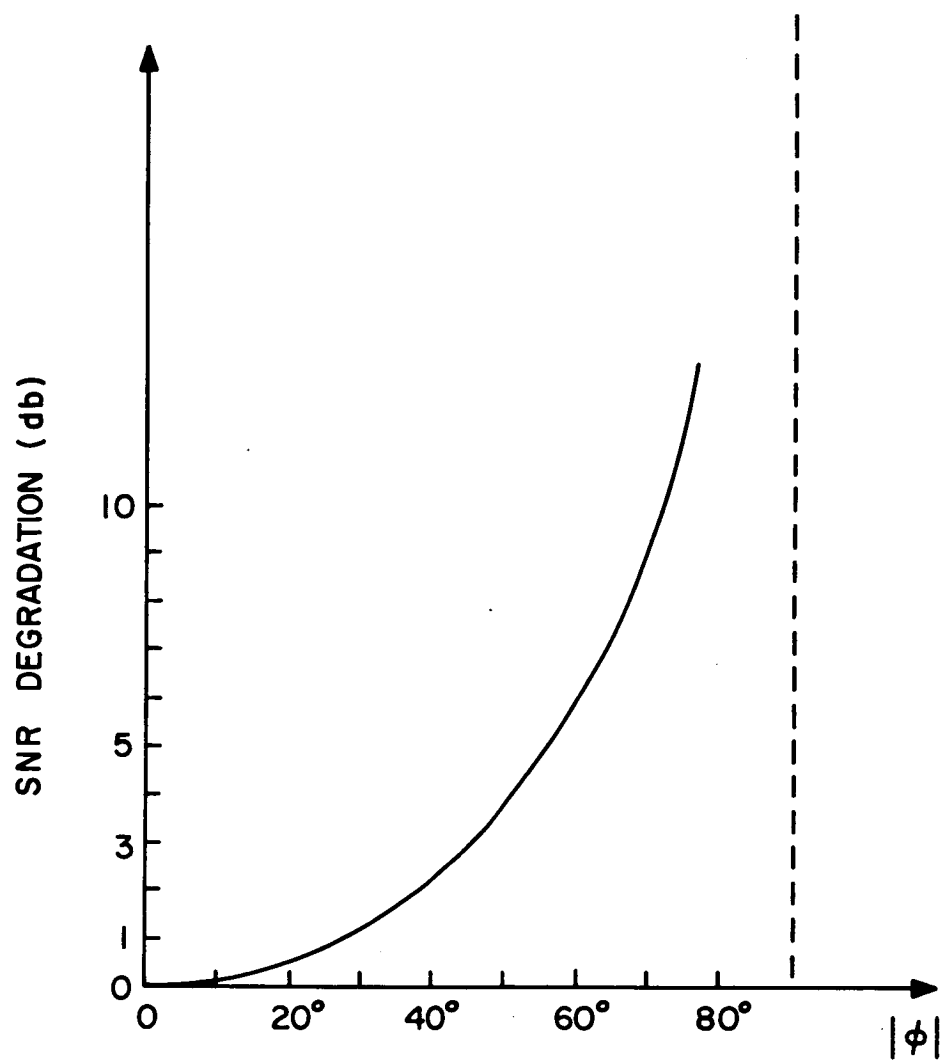


Fig. 7

SNR degradation due to constant phase error

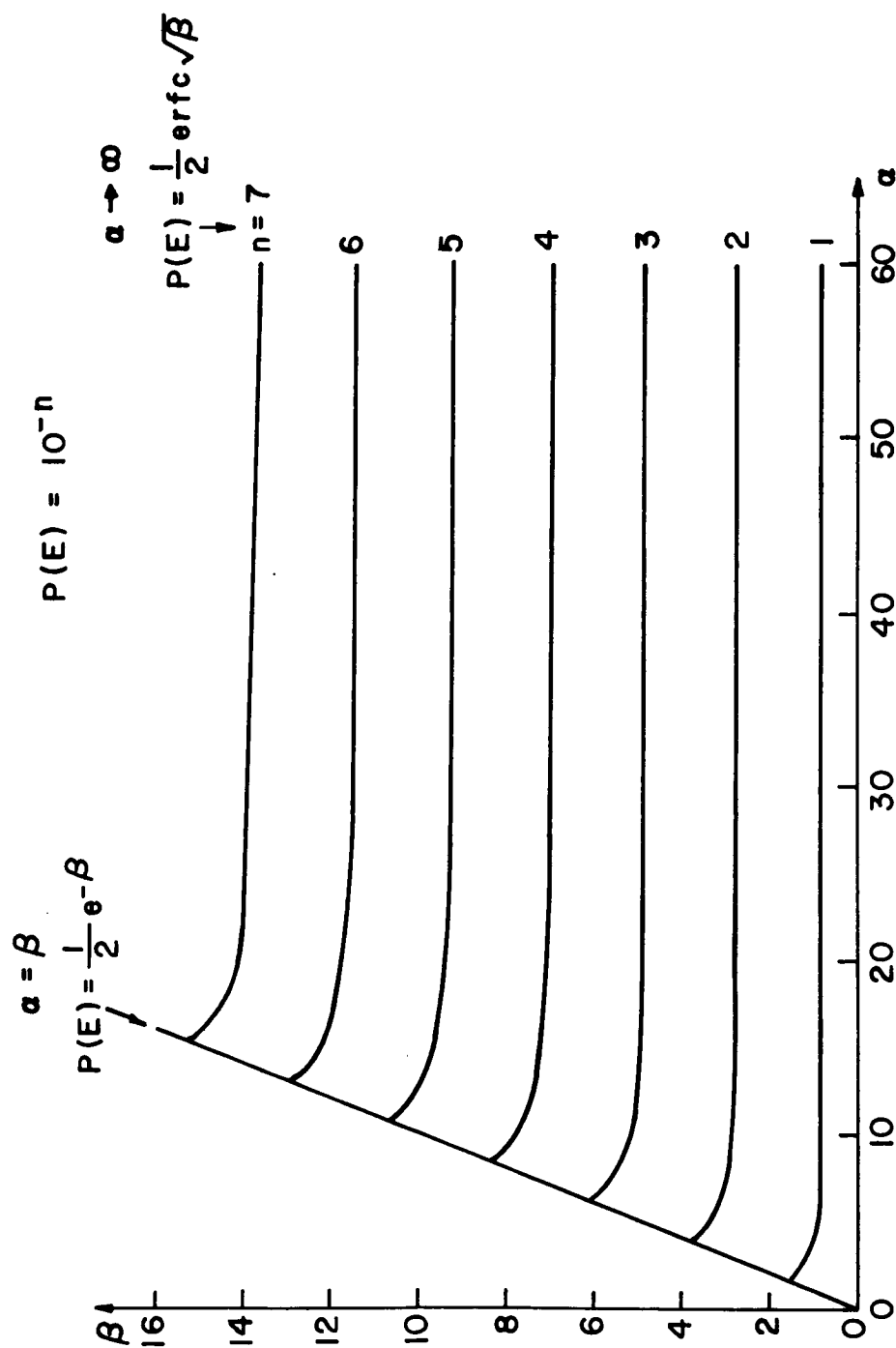


FIG. 8

SNR degradation due to constant phase error

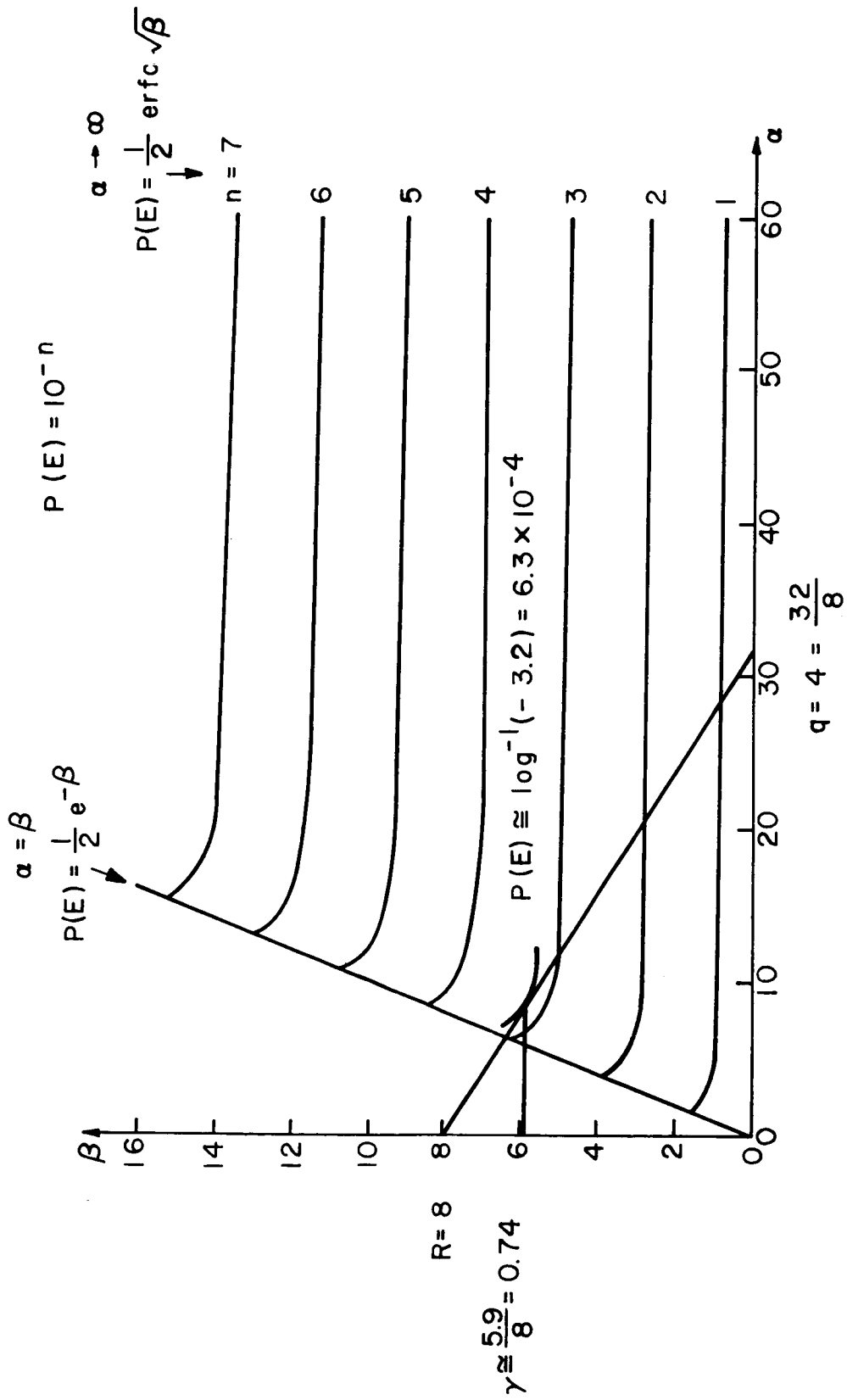


FIG. 9 Design Example

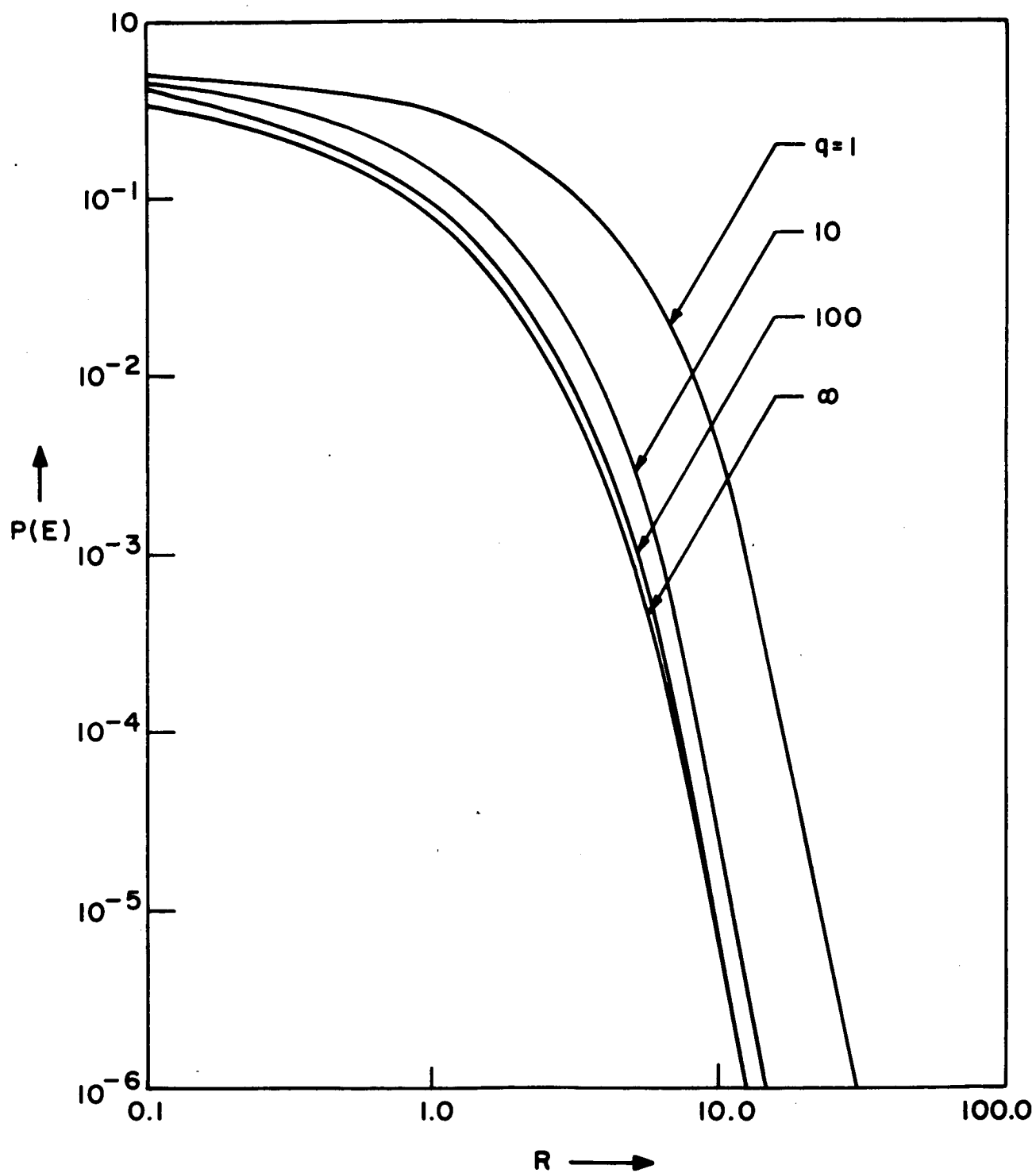


Fig. 10

Minimum $P(E)$ vs. R (parameter q)

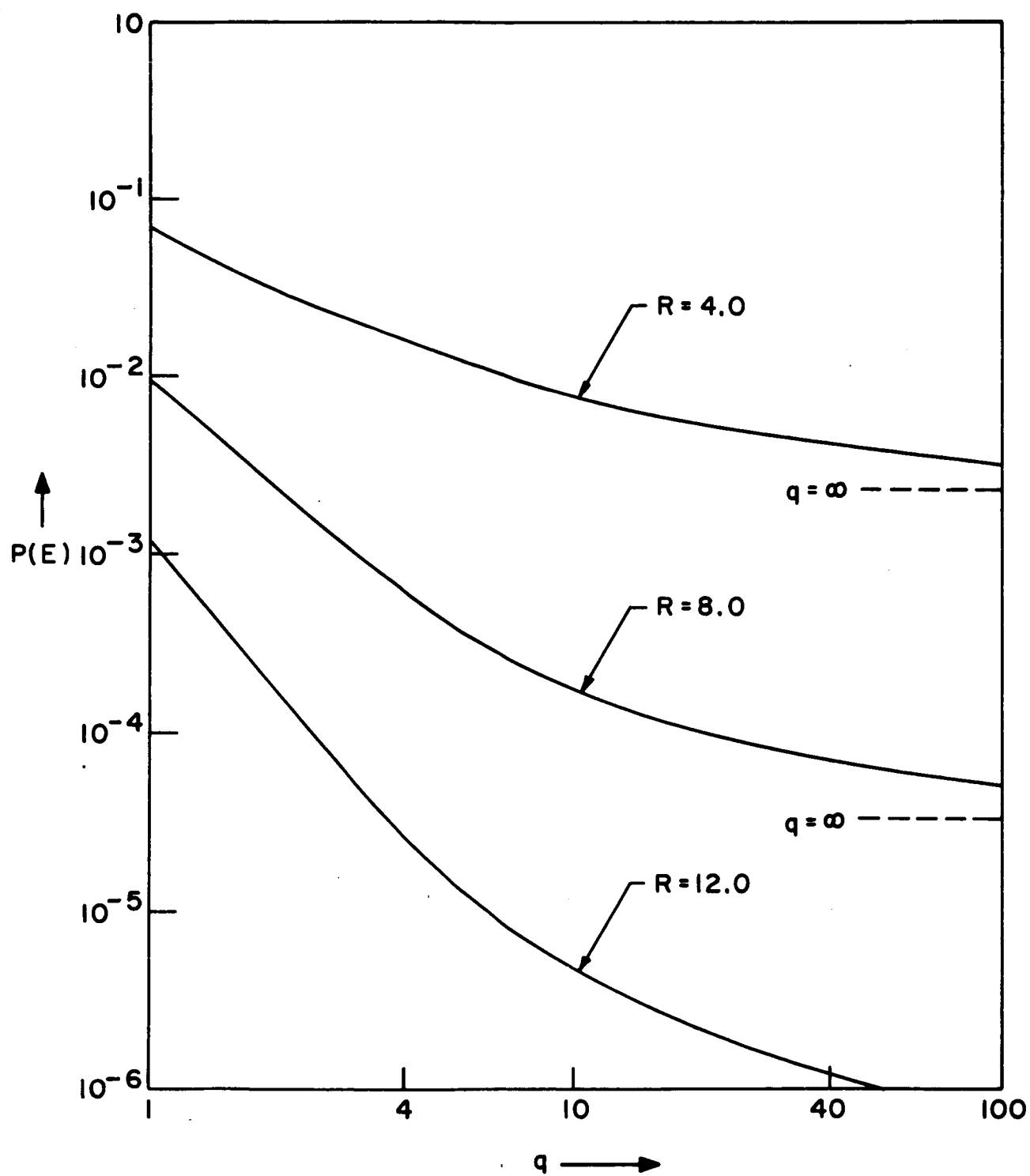


Fig. 11

Minimum $P(E)$ vs. q (parameter R)

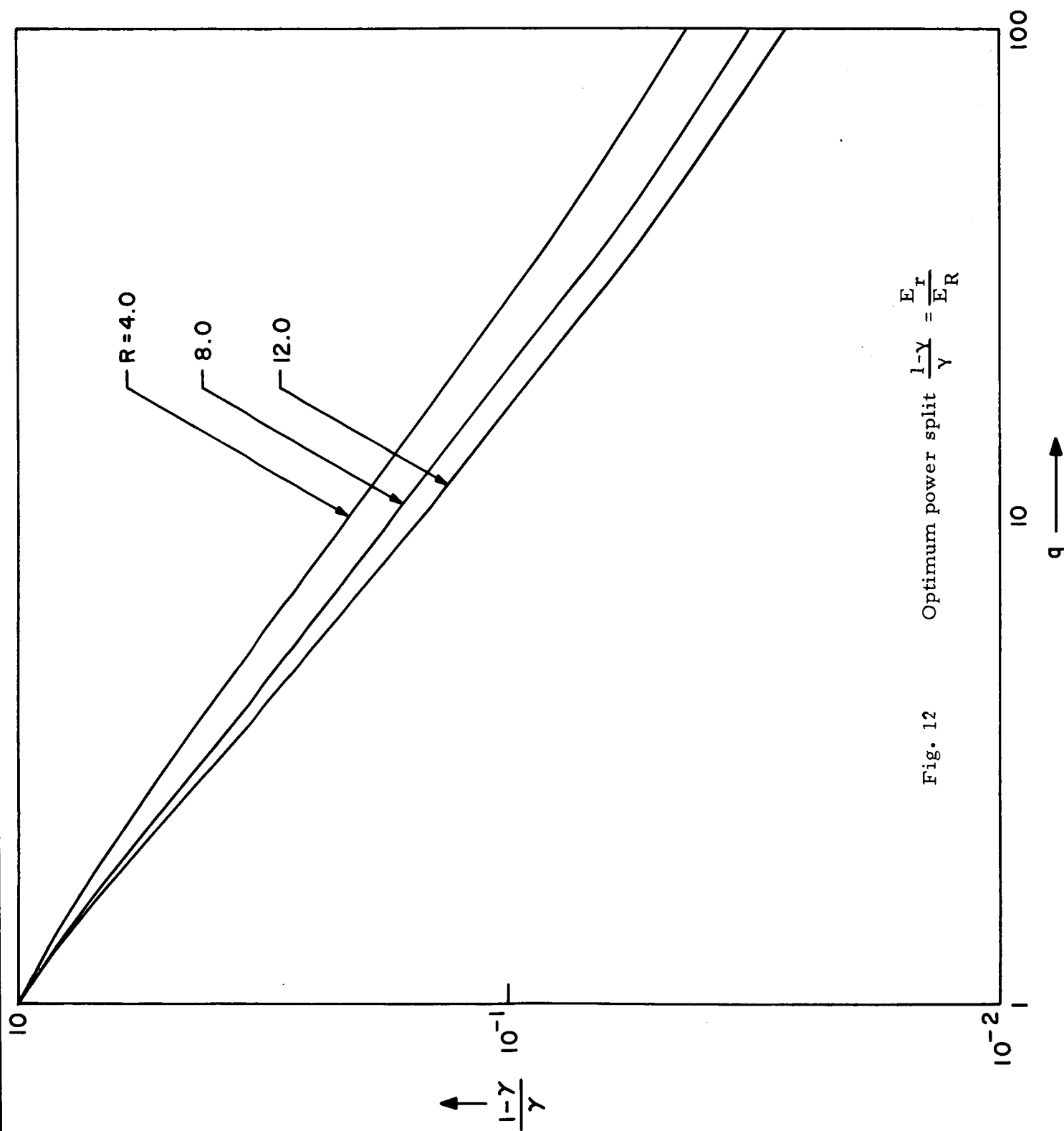


Fig. 12 Optimum power split $\frac{1-\gamma}{\gamma} = \frac{E_r}{E_R}$

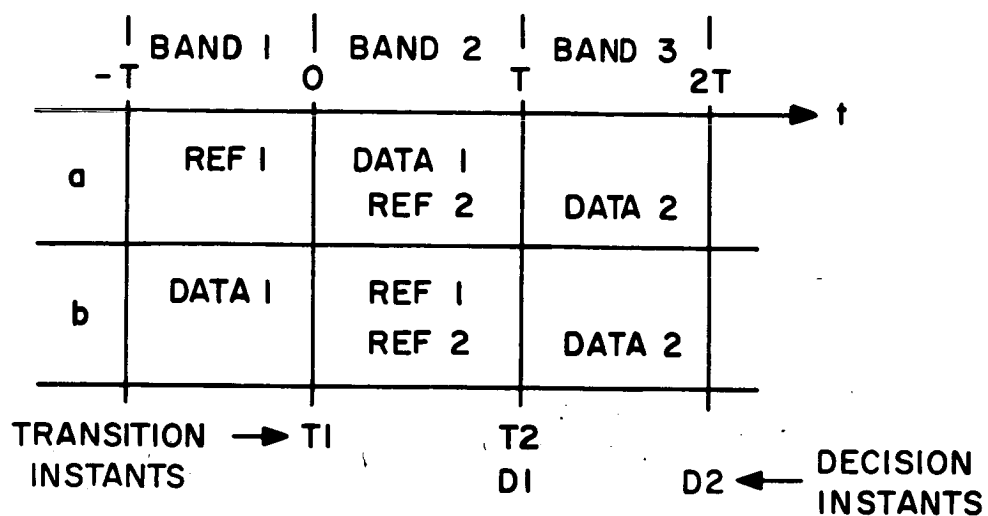


Fig. 13 Signal bands for D-PSK

STATE	DETECTION		DECODING	
	PRESENT	PRECEDING	ABSOLUTE	DIFFERENTIAL
1	CORRECT	CORRECT	CORRECT	CORRECT
2	CORRECT	INCORRECT	CORRECT	INCORRECT
3	INCORRECT	CORRECT	INCORRECT	INCORRECT
4	INCORRECT	INCORRECT	INCORRECT	CORRECT

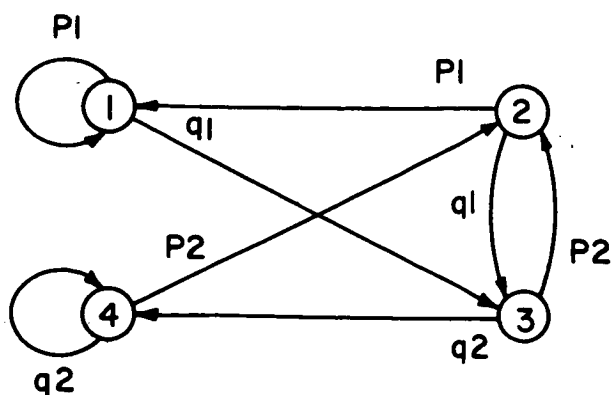


Fig. 14 TR system state diagram

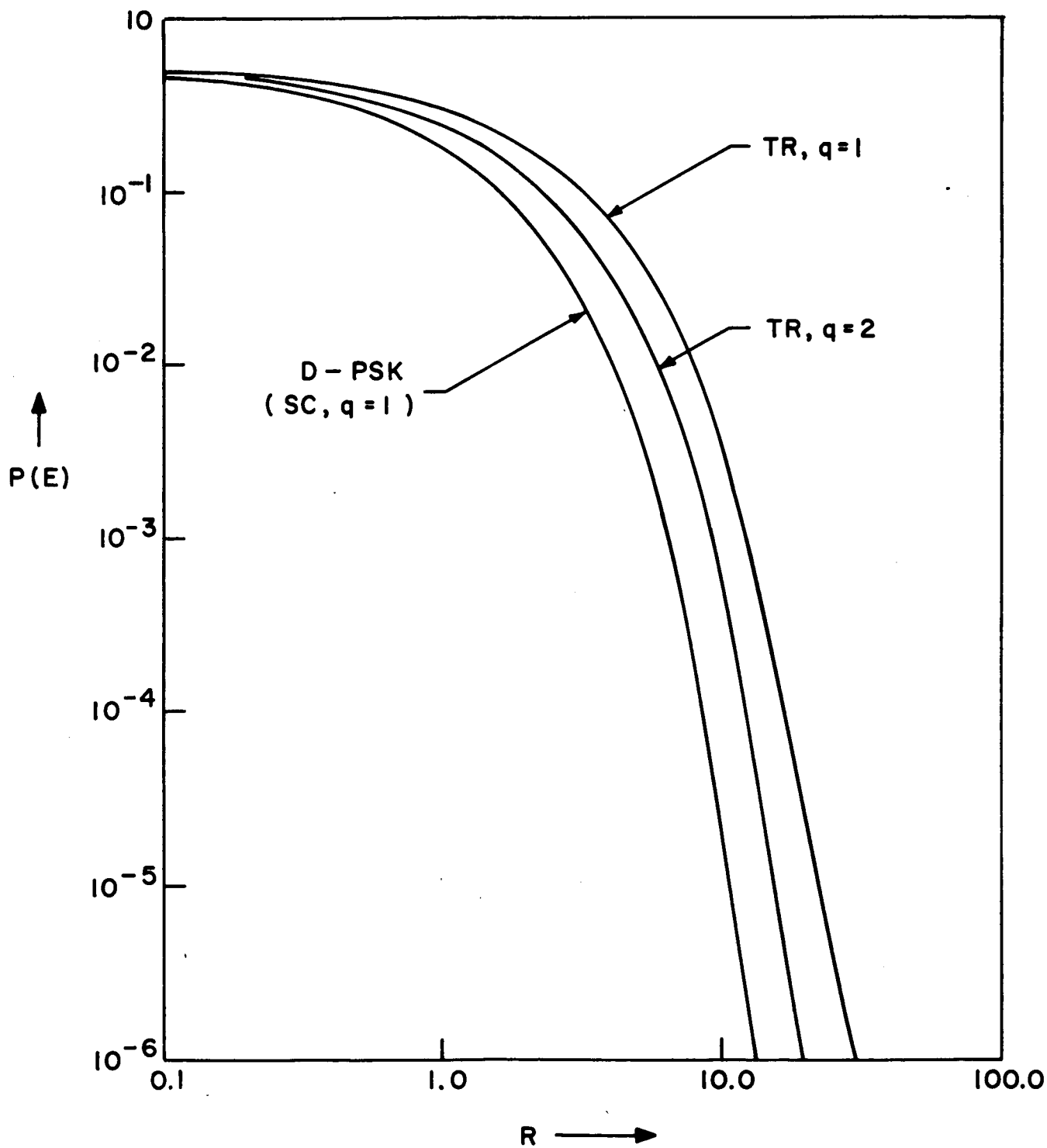


Fig. 15

Comparison of TR and SC systems for small q

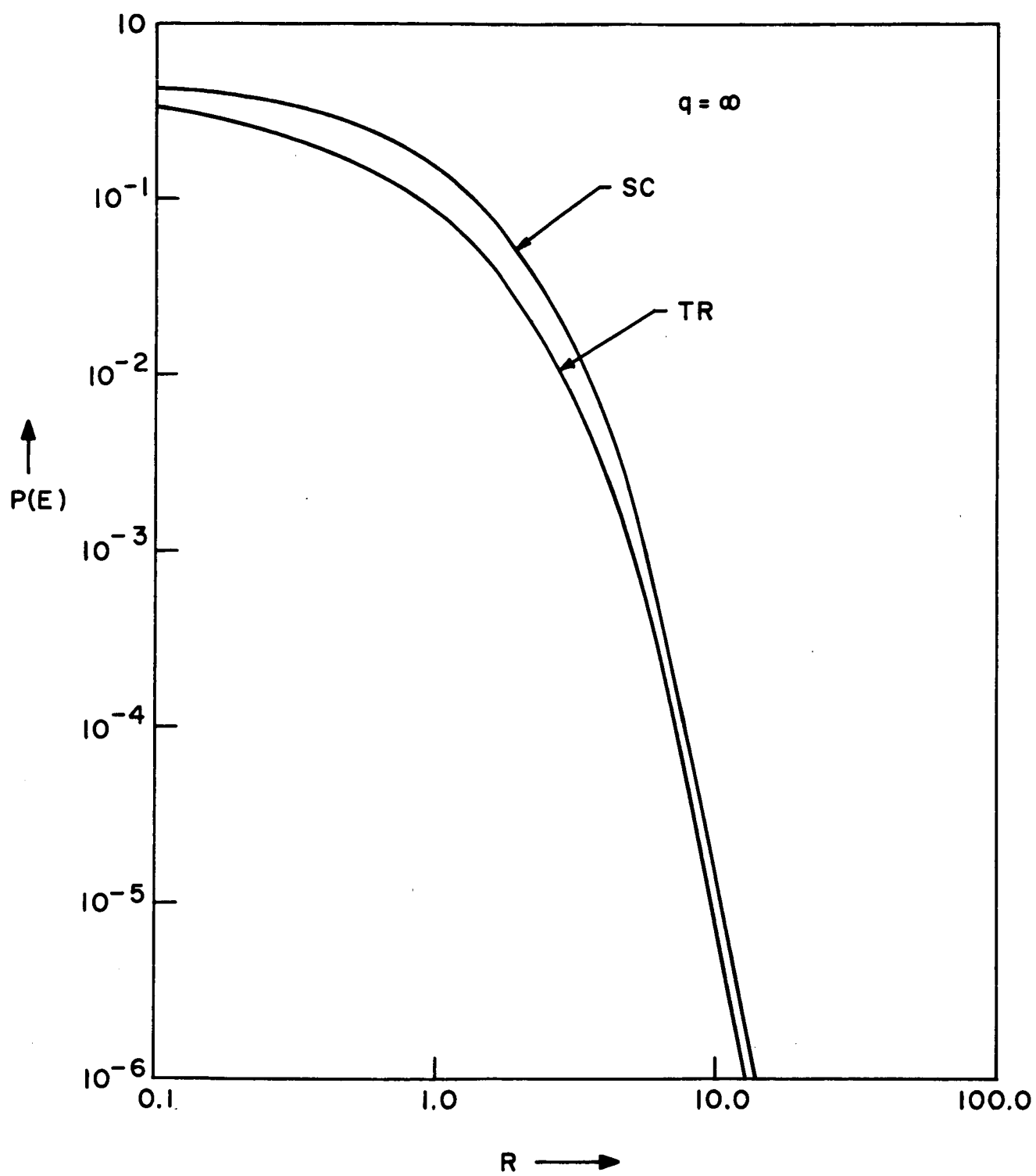


Fig. 16

Comparison of TR and SC systems for large q

IV. - Recursive Detection

I. Introduction

The purpose of this work is to develop a different formulation for the classical problem of detecting the presence of waveform signals in additive, Gaussian, non-white, noise. Unlike the usual formulation of this problem^{1, 2} the present treatment does not involve the use of ortho-normal expansions of the covariance kernel of the noise, nor does the result involve an integral equation which is tedious to solve.³ In this paper, for the discrete time case, which is important in digital processing, the test statistic is generated recursively by means of a first order difference equation. This results in minimizing the memory requirements and programming effort. Furthermore, the highest matrix inversion is dependent only on the statistics of the noise and not on the number of samples taken. This simplification is made possible by assuming that the additive noise is generated as the solution of a linear differential equation driven by white noise.⁴ This manner of generating the noise is natural in physical problems and for the stationary case is identical to specifying the spectral density. An appropriate state vector of this solution is then Markov and can therefore be completely characterized by its transition (conditional) probability density functions. Because of the nature of the formulation, it is just as easy to establish the test statistic when the noise is non-stationary provided that the time-varying differential equation generating the noise is known. Furthermore, if the differential equation is not linear, the Markov structure is retained, and although the noise is not Gaussian, likelihood ratios may still be found. This is not pursued in this paper.

In section II, the difference equation for the likelihood ratio together with initial conditions are derived. The result is exhibited directly in terms of the coefficients of the noise-generating differential equation and the signal and several of its derivatives. A similar derivation is presented in section III

for non-stationary noise and explicit results are presented as an example involving a first order time-varying system.

By allowing the discrete sample spacing to reduce to zero, the difference equations derived in sections II and III are shown in section IV to yield a first order differential equation. The likelihood can therefore be written explicitly since the initial conditions are known. The form of the solution is interpreted as the usual prewhitening-matched filter combination.⁵ In addition, the initial conditions are derived. These initial conditions may, in some instances, contribute significantly to minimizing the probability of error. The signal to noise ratio is calculated for the general cases of both stationary and non-stationary noise. This, in itself, is an extremely useful result in that the performance (probability of error) of the optimum receiver for a very broad class of noise statistics may be obtained directly. Furthermore, since the expression for the signal-to-noise ratio is so compact, it lends itself readily to such applications as signal design and effect of variations of parameters.

Section V treats the various extensions of the basic formulation including the problem where the noise is generated by a differential equation driven by white noise and its derivative (equivalent to a spectral density with a numerator, in the stationary case) and possible application to signal design, diversity, composite hypotheses, etc.

The idea of deriving a recursive equation for the likelihood ratio was first presented by Schweppe⁶ who discusses the detection of random signals in white noise. This was applied to the detection of known signals in first-order Markov noise by Kleinberg.⁷ Some of the results in this work have also been derived with the use of a Reproducing Kernel Hilbert Space^{8, 9, 10, 11}. The results of this research was presented at the 1967 International Symposium on Information Theory in San Reno, Italy. Details of this work are contained in a paper submitted to the Transactions on Information Theory of the IEEE.¹⁴

II. Problem Formulation - Stationary Noise

It is assumed that one of two completely known signals $s_1(t)$ or $s_2(t)$ is received in additive, zero-mean Gaussian noise. The received signal is then

$$Y(t) = s_i(t) + X(t), \quad i = 1, 2, \quad 0 \leq t \quad (1)$$

The noise process $X(t)$ is assumed to be generated as the solution of a k^{th} order differential equation driven by zero-mean white noise. That is,

$$L[X(t)] = \left(\frac{d^k}{dt^k} + \dots + a_{n-1} \frac{d^{k-1}}{dt^{k-1}} + \dots + a_1 \frac{d}{dt} + a_0 \right) X(t) = W(t) \quad (2)$$

where $L[]$ is the linear differential operator described above and $W(t)$ is Gaussian white noise with spectral density n_0 . (This representation of the noise is equivalent, with constant coefficients a_i , to considering stationary noise with a rational spectral density and a constant numerator. The extension to non-constant numerators will be discussed later.)

Equation (12) can be written in matrix form as

$$\dot{\underline{X}}(t) = A \underline{X}(t) + \underline{b} W(t) \quad (3)$$

where $\underline{X}(t)$ is an appropriate state vector for the system. For this section the following natural choice can be made.

$$\underline{X}(t) = \begin{bmatrix} X(t) \\ \dot{X}(t) \\ \vdots \\ X^{(k-1)}(t) \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -a_0 & -a_1 & -a_2 & -a_3 & \dots & -a_{k-1} \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Here $\underline{X}(t)$ and \underline{b} are k -dimensional column vectors and A is a $k \times k$ matrix.

The noise vector process $\underline{X}(t)$ is thus Markov as well as Gaussian, and is completely characterized by its conditional mean vector and conditional covariance matrix.

The receiver that minimizes the probability of error, in deciding^{1, 2, 3} between the two signals, is usually found by considering the likelihood ratio of the received random waveform $Y(t)$. To make use of the Markovian nature of the noise, consider the likelihood ratio of the vector process $\underline{Y}(t)$ (defined in the same manner as $\underline{X}(t)$). Since $Y(t)$, which is sufficient for optimum detection, is contained in the vector $\underline{Y}(t)$, the resultant receiver will still be optimum.

In order to write the likelihood ratio in a convenient form the process $\underline{Y}(t)$ will be considered to be sampled at equally spaced intervals of duration Δ . (For digital processing, Δ will be the appropriate spacing and the digital nature of the receiver will be as described below. For analog processing Δ will be made arbitrarily small, and a limiting form for the receiver will be derived). Then, considering only the first t seconds of data, where $t = n \Delta$, and using the notation $\underline{Y}_k = \underline{Y}(k\Delta)$, the likelihood ratio is

$$L(\underline{Y}_1, \dots, \underline{Y}_n) = \frac{f_{\underline{Y}_1, \dots, \underline{Y}_n}(\underline{Y}_1, \dots, \underline{Y}_n | S_2)}{f_{\underline{Y}_1, \dots, \underline{Y}_n}(\underline{Y}_1, \dots, \underline{Y}_n | S_1)}$$

$$= \frac{f_{\underline{Y}_n}(\underline{Y}_n | \underline{Y}_{n-1}, S_2) f_{\underline{Y}_1, \dots, \underline{Y}_{n-1}}(\underline{Y}_1, \dots, \underline{Y}_{n-1} | S_2)}{f_{\underline{Y}_n}(\underline{Y}_n | \underline{Y}_{n-1}, S_1) f_{\underline{Y}_1, \dots, \underline{Y}_{n-1}}(\underline{Y}_1, \dots, \underline{Y}_{n-1} | S_1)}$$

where the conditioning S_i refers to the signal $s_i(t)$ having been transmitted and the Markov properties of $\underline{Y}(t)$ have been used. Taking logarithms, and defining the logarithm of the likelihood ratio of the first $t = n \Delta$ seconds of data as θ_n , we obtain

$$\theta_n - \theta_{n-1} = \ln \frac{f_{\underline{Y}_n}(\underline{Y}_n | \underline{Y}_{n-1}, S_2)}{f_{\underline{Y}_n}(\underline{Y}_n | \underline{Y}_{n-1}, S_1)} \quad (4)$$

Using equation (1), assuming that the noise $X(t)$ is independent of which signal was transmitted, and defining a signal vector $\underline{s}_i(t)$ in a manner similar to the vector random processes $\underline{X}(t)$ and $\underline{Y}(t)$, equation (4) becomes

$$\theta_n - \theta_{n-1} = \ln \frac{f_{\underline{X}_n}(\underline{Y}_n - \underline{s}_{2,n} | \underline{X}_{n-1} = \underline{Y}_{n-1} - \underline{s}_{2,n-1})}{f_{\underline{X}_n}(\underline{Y}_n - \underline{s}_{1,n} | \underline{X}_{n-1} = \underline{Y}_{n-1} - \underline{s}_{1,n-1})} \quad (5)$$

Now, since \underline{X}_n and \underline{X}_{n-1} are jointly normal,¹²

$$f_{\underline{X}_n}(\underline{x}_n | \underline{x}_{n-1}, S_i) = \frac{1}{(2\pi)^{k/2} |K|^{1/2}} \exp \left[-\frac{1}{2} (\underline{x}_n - \underline{\mu}_i)^T K^{-1} (\underline{x}_n - \underline{\mu}_i) \right]$$

where $\underline{\mu}_i$ and K are the mean vector and covariance matrix, respectively, of the random vector \underline{X}_n , conditioned upon $\underline{X}_{n-1} = \underline{x}_{n-1}$. Hence equation (5) simplifies to

$$\theta_n - \theta_{n-1} = \frac{1}{2} (a_2 - a_1) \quad (6)$$

where

$$a_i = [(\underline{Y}_n - \underline{s}_{i,n}) - \underline{\mu}_i]^T K^{-1} [(\underline{Y}_n - \underline{s}_{i,n}) - \underline{\mu}_i], \quad i = 1, 2.$$

The initial conditions for the difference equation (6) are given by the logarithm of the likelihood ratio for the single set of samples taken at $t = 0$,

$$\theta_0 = \ln \frac{f_{\underline{X}_0}(\underline{Y}_0 - \underline{s}_{2,0})}{f_{\underline{X}_0}(\underline{Y}_0 - \underline{s}_{1,0})} \quad (7)$$

In reference (14) it is shown that $\underline{\mu}_i$ and K can be written in terms of $\underline{X}_{n-1} = \underline{Y}_{n-1} - \underline{s}_{i, n-1}$ and the unconditional covariance matrix

$$C_{\Delta} \equiv E \underline{X}_{n-1} \underline{X}_{n-1}^T = C(\Delta) = E \underline{X}(t) \underline{X}(t-\Delta).$$

Then for stationary noise,

$$\underline{\mu}_i = C_{\Delta} C_o^{-1} \underline{X}_{n-1}, \quad (8)$$

and

$$K = C_o - C_{\Delta} C_o^{-1} C_{\Delta}^T. \quad (9)$$

Since $C(t) = e^{At} C_o \quad (14),$

equations (8) and (9) now simplify to

$$\underline{\mu}_i = e^{A\Delta} (\underline{Y}_{n-1} - \underline{s}_{i, n-1}), \quad (10)$$

and

$$K = C_o - e^{A\Delta} C_o e^{A^T \Delta}. \quad (11)$$

Furthermore equation (7) can be written as

$$\theta_o = -\frac{1}{2} (\underline{Y}_o - \underline{s}_2, o)^T C_o^{-1} (\underline{Y}_o - \underline{s}_2, o) + \frac{1}{2} (\underline{Y}_o - \underline{s}_1, o)^T C_o^{-1} (\underline{Y}_o - \underline{s}_1, o). \quad (12)$$

The matrix C_o can now be found to satisfy the following matrix equation

$$A C_o + C_o A^T = -n_o \underline{b} \underline{b}^T. \quad (13)$$

In the above, $\underline{b}\underline{b}^T$ is an extremely simple matrix, having all zero elements, except for unity in the lower right hand corner. Equations (6), (10), (11), (12) and (13) now suffice to determine the recursion formula for the test statistic. (An expression for the elements of the matrix C_0 directly in terms of the elements a_i of A will be derived shortly).

Example: Second order case ($k = 2$).

Here $A = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix}$. Equation (13) becomes

$$\begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} 0 & -a_0 \\ 1 & -a_1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -n_0 \end{bmatrix},$$

which easily yields $c_{12} = c_{21} = 0$, $c_{22} = \frac{n_0}{2a_1}$, $c_{11} = \frac{n_0}{2a_1 a_0}$. These can easily be verified since, from equation (2), the covariance function of $X(t)$ can be found and $c_{11} = EX^2(t)$, $c_{12} = C_{21} = EX(t) \dot{X}(t) = 0$, $c_{22} = E\dot{X}^2(t)$. There are standard techniques for finding $e^{A\Delta}$.¹³

III. Non-stationary Noise

If the coefficients of equation (2) and/or the strength of the white noise $W(t)$ are time-varying, then the additive noise $X(t)$ is non-stationary and the mean and the covariance matrix given in equations (8) and (9) must be modified. From the appendix,

$$\underline{\mu}_i = C_{n, n-1} C_{n-1, n-1}^{-1} \underline{X}_{n-1}, \quad (14)$$

and

$$K_n = C_{n, n} - C_{n, n-1} C_{n-1, n-1}^{-1} C_{n, n-1}^T \quad (15)$$

where

$$C_{n, k} = E \underline{X}_n \underline{X}_k^T = E \underline{X}(n\Delta) \underline{X}(k\Delta)^T = C(n\Delta, k\Delta), \text{ and } K_n$$

is to replace K in the above expressions.

The symmetric covariance matrix of the noise, $C(t, t)$, can be found⁽¹⁴⁾ from

$$\frac{dC(t, t)}{dt} = A(t) C(t, t) + C(t, t) A^T(t) + n_o(t) \underline{b} \underline{b}^T, \quad (16)$$

where $\Phi(t, u)$, the fundamental matrix of (3), satisfies the differential equation

$$\frac{\partial \Phi(t, u)}{\partial t} = A(t) \Phi(t, u), \quad \Phi(u, u) = I.$$

Equation (16) can be solved to yield

$$C(t, t) = \int_{-\infty}^t n_o(u) \Phi(t, u) \underline{b} \underline{b}^T \Phi^T(t, u) du \quad (17)$$

which defines $C_{n, n}$ or $C_{n-1, n-1}$. Furthermore, it can be shown that

$$C(t, u) = \Phi(t, u) C(u, u), \quad t \geq u. \quad (18)$$

This last expression can be used to find $C_{n, n-1}$.

It is not possible to go further in the time-varying case without knowledge of the fundamental matrix $\Phi(t, u)$. One important situation where an explicit solution is possible, is when the noise $\underline{X}(t)$ is generated by a first order time-varying differential equation. That is,

$$\dot{X}(t) = a(t) X(t) + W(t)$$

The solution is scalar Markov so that all matrices become scalars. The fundamental matrix (in this case a scalar) is now

$$\phi(t, u) = e^{\int_u^t a(v) dv}.$$

Then

$$C(t, t) = EX^2(t) = \int_{-\infty}^t n_o(u) \phi^2(t, u) du,$$

and

$$C(t, u) = \phi(t, u) \int_{-\infty}^t n_o(v) \phi^2(t, v) dv, \quad t \geq u.$$

Since these are scalars, we can now write

$$\mu_i = \frac{C(n\Delta, (n-1)\Delta)}{C((n-1)\Delta, (n-1)\Delta)} X_{n-1},$$

and

$$K_n = C(n\Delta, n\Delta) - \frac{C(n\Delta, (n-1)\Delta)^2}{C((n-1)\Delta, (n-1)\Delta)}$$

The difference equation (6) can now be written easily.

IV. Analog Processing

We consider here the results of the last two sections when the spacing Δ is made arbitrarily small. The resultant optimum receiver (i.e., likelihood ratio) should then be identical to that obtained using conventional Karhunen-Loève techniques. If we assume that $s_2(t) = -s_1(t) = s(t)$ for simplicity (if $s_2(t) \neq -s_1(t)$, then a similarly derived bias term must be added to the likelihood ratio), then equation (6) simplifies to

$$\dot{\theta}(t) = \frac{2}{n_0} L[Y(t)] L[s(t)]. \quad (19)$$

The initial conditions for equation (19) are derived from equation (12), which simplifies to

$$\theta(0) = 2 \underline{Y}(0)^T C_0^{-1} \underline{s}(0). \quad (20)$$

Hence,

$$\theta(t) = \frac{2}{n_0} \int_0^t L[Y(u)] L[s(u)] du + 2 \underline{Y}(0)^T C_0^{-1} \underline{s}(0), \quad (21)$$

the likelihood ratio for stationary noise.

The coefficients c^{ij} of C_0^{-1} are easily found^(11, 14) to be

$$c^{ij} = \begin{cases} \sum_{l=\max(0, i+j-1-k)}^{\min(i, j)} (-1)^{i-l} a_l a_{2l+k-i-j-1}, & i+j \text{ even} \\ 0 & i+j \text{ odd} \end{cases}$$

Equation (26) can be modified to resemble the form of the solution when Karhunen-Loève techniques are used.⁽¹⁴⁾ This will reduce the number of data integrations required. The first term in equation (26) could also be found by considering a prewhitening filter.⁵ If $X(t)$ is passed through a filter with transfer function

$$H(\omega) = \sum_{j=0}^k a_j (i\omega)^j,$$

(note that this is an instantaneous device and does not spread the incoming waveform), then the output is white and the first part of the term in equation (26) is the matched filter or correlator solution. We have here provided a method to obtain the initial conditions, without which the solution is invalid. The signal used in the matched filter or correlator is obtained by passing the original signal through the same filter $H(\omega)$. Equivalent to $H(\omega)$ in the time domain is the linear differential operator $L[]$ defined in equation (2).

An alternate method of implementing the receiver is to use, as a signal in the matched filter or correlator, the output of the filter $|H(\omega)|^2$ which is driven by the original signal. In the time domain, this corresponds to the $2k^{\text{th}}$ order linear differential operator L^+L , while the operator L^+ corresponds to the transfer function $H^*(\omega)$. We have again supplied the terms missing from the prewhitening solution.

An explicit general expression for the optimum signal-to-noise ratio for the optimum signal-to-noise ratio for stationary statistics is found to be

$$\gamma_\theta = \frac{1}{n_0} \int_0^t L[s(u)]^2 du + \underline{s}(0)^T C_0^{-1} \underline{s}(0) \quad (14) \quad (22)$$

For any signal $s(t)$ and with C_0 computed from equation (13) the signal-to-noise ratio, hence the probability of error can easily be found. It is now possible to design signals by maximizing γ_0 subject to constraints such as energy, bandwidth, and intersymbol interference. The first term in equation (22) is the signal-to-noise ratio for the pre-whitening filter. Thus the last term gives the improvement in using initial conditions.

As mentioned above, an equation similar to (26) can also be derived using Karhunen-Loève techniques. However the general and compact form of our result has not been obtained because of the tedium of the computations necessary with that method. Our technique avoids those difficulties.

The above derivation was for stationary noise. The extension to non-stationary noise is not very difficult. Indeed, we get

$$\dot{\theta}(t) = \frac{2}{n_0(t)} L_t[Y(t)] L_t[s(t)] \quad (23)$$

where the subscript t denotes a time-varying differential operator. The initial conditions are now

$$\dot{\theta}(0) = 2 \underline{Y}(0)^T C(0, 0)^{-1} \underline{s}(0)$$

where $C(0, 0)$ can be found from equation (17). Combining these two we have,

$$\theta(t) = 2 \int_0^t \frac{L_u[Y(u)] L_u[s(u)]}{n_0(u)} du + 2 \underline{Y}(0)^T C(0, 0)^{-1} \underline{s}(0) \quad (24)$$

Integrating by parts will again reduce the number of derivatives of $Y(t)$ that must be taken. For the example at the end of section III,

$$C(0, 0) = \int_{-\infty}^0 n_0(u) \phi^2(0, u), \text{ a scalar.}$$

In addition to providing the detector for non-stationary noise, equation (29) can be used to find the resultant maximum signal-to-noise

ratio. In a manner, similar to that for non-stationary noise, we easily find that

$$\gamma_{\theta} = \int_0^t \frac{L_u[s(u)]^2}{n_o(u)} du + \underline{s}(0)^T C(0, 0)^{-1} \underline{s}(0) \quad (25)$$

The probability of error can now be computed. For example, in the first order case discussed above,

$$\gamma_{\theta} = \int_0^t \frac{[\dot{s}(u) - a(u)s(u)]^2}{n_o(u)} du + \frac{s^2(0)}{C(0,0)},$$

where $C(0, 0)$ has already been described.

In many non-stationary problems only the noise power strength $n_o(t)$ is time-varying while the differential equation is not. Thus $A(t)$ is a constant matrix, the fundamental matrix $\Phi(t, u) = e^{A(t-u)}$, as in the stationary case, and the equation for $C(t, t)$ simplifies to

$$C(t, t) = \int_{-\infty}^t n_o(u) e^{A(t-u)} \underline{b} \underline{b}^T e^{A^T(t-u)} du,$$

which is easy to solve. Thus, for this case, the expressions (29) and (30) for the optimum detector and signal-to-noise ratio, respectively, become explicit functions of system parameters and yield convenient solutions for this non-stationary problem.

V. Conclusions and Extensions

In conclusion, we have developed a recursive technique for finding a difference equation for the likelihood ratio. This likelihood ratio describes the optimum detector for deciding between one of two known signals in the presence of additive non-white Gaussian noise. We have considered the noise as being generated by a system which is described by a linear time-varying differential equation driven in turn by white non-stationary noise. (The case when the system includes derivatives of the input noise is discussed below.) For digital processing we have described the difference equation and discussed the evaluation of its components.

For analog processing we have been able to derive, by a suitable limiting argument, the likelihood ratio in a compact and general form. This form has some similarity to that derived by prewhitening arguments, but contains additional terms which are neglected by that method. This result can also be obtained by standard Karhunen-Loève techniques which involve tedious computations for the terms involving initial and final values of the variables (see equation (27)). Furthermore, we have been able to derive compact and general expressions for the signal-to-noise ratios for both stationary and non-stationary statistics. This result easily allows calculation of probability of error and comparison with sub-optimum detectors, such as the prewhitening filter.

In the general case of non-white noise, the right hand side of the differential equation defining the noise contains derivatives of the white process. Then we have

$$\begin{aligned}
 L[X(t)] &= \left(\frac{d^k}{dt^k} + a_{k-1} \frac{d^{k-1}}{dt^{k-1}} + \dots + a_1 \frac{d}{dt} + a_0 \right) X(t) \\
 &= \left(\frac{d^\ell}{dt^\ell} + b_{\ell-1} \frac{d^{\ell-1}}{dt^{\ell-1}} + \dots + b_1 \frac{d}{dt} + b_0 \right) W(t) \\
 &= M[W(t)]
 \end{aligned} \tag{31}$$

Here L and M are differential operators of degree k and l respectively, with $k > l$. It is then always possible¹³ to define a state vector $\underline{X}(t)$ for this system such that equation (31) simplifies to

$$\dot{\underline{X}}(t) = A \underline{X}(t) + \underline{b} W(t) \quad (32)$$

where \underline{b} is a $1 \times k$ column vector with elements b_j . The derivation of the difference equation for the likelihood ratio can now proceed. It can then be shown that, for stationary statistics,

$$\dot{\theta}(t) = \frac{u(t) v(t)}{n_o} \quad (33)$$

where $u(t)$ and $v(t)$ are solutions of the differential equations

$$\begin{aligned} M[u(t)] &= L[s(t)] , \\ M[v(t)] &= L[Y(t)] , \end{aligned}$$

with appropriate initial conditions. Similar results can be derived for the non-stationary case.

Since the origin of noise in many physical situations is identical to the model used in this paper, it is worthwhile to pursue generalizations of these methods. A natural extension is the processing of several correlated signals as in diversity. The noise would then be generated by a vector equation similar to equation (3) with the dimensionality multiplied by the order of the diversity. Another extension is to the problem of composite hypothesis testing as for incoherent detection. Finally, the possibility of taking advantage of the Markov property of the noise even if it is not Gaussian may be considered. These and other problems such as signal design are being investigated.

REFERENCES

1. Helstrom, C. W. Statistical Theory of Signal Detection, Pergamon Press, 1960.
2. Schwartz, M., W. R. Bennett, and S. Stein. Communication Systems and Techniques, McGraw-Hill Book Co., 1966.
3. Davenport, W. B., And W. L. Root. An Introduction to the Theory of Random Signals and Noise, McGraw-Hill Book Co., 1958.
4. Kalman, R. E. "A new Approach to Linear Filtering and Prediction Problems," J. Basic Eng. (ASME Trans.) 82D, 35, March, 1960.
5. Selin, J. Detection Theory, Princeton University Press, 1965.
6. Scheppe, F. C. "Evaluation of Likelihood Functions for Gaussian Signals", IEEE Trans. on Information Theory, Vol. IT-11, pp. 61-70, Jan. 1965.
7. Kleinberg, L. "Evaluation of Likelihood Ratios for Deterministic Signals in Gaussian Noise," M. S. (E. E.) Report, Polytechnic Institute of Brooklyn, June, 1967.
8. Parzen, E. "Extraction and Detection Problems and Reproducing Kernel Hilbert Spaces, Journal SIAM Control, Vol. 1, 1962, pp. 35-62.
9. Kailath, T. Statistical Detection Theory (to be published by Prentice Hall)
10. Capon, J. "Hilbert Space Methods for Detection Theory and Pattern Recognition," IEEE Trans. on Information Theory, Vol. IT-11, pp. 247-259, April 1965.
11. Hajek, J. "On Linear Statistical Problems in Stochastic Processes," Czechoslovak Math. Journ., Vol. 87, 1962, pp. 404-444.
12. Papoulis, A. Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Co., 1965.
13. Zadeh, L. A., and C. A. Desoer. Linear System Theory, McGraw-Hill Book Co., 1963.
14. Pickholtz, R., and R. Boorstyn. "A Recussive Approach to Signal Detection", presented at 1967 International Symposium on Information Theory in San Remo, Italy and submitted to the IEEE Trans. on Information Theory.

Appendix

The conditional mean vector of a normal vector random variable \underline{X}_n is a linear function of the conditioning vector \underline{X}_{n-1} if \underline{X}_n and \underline{X}_{n-1} are jointly normal. Hence, assuming zero means

$$E(\underline{X}_n | \underline{X}_{n-1}) = B \underline{X}_{n-1}$$

where B is a $k \times k$ matrix. Furthermore, since the noncentral second moment is minimized when taken about the mean,¹² B must minimize

$$\epsilon^2(\underline{X}_{n-1}) \equiv E[(\underline{X}_n - B \underline{X}_{n-1})^T (\underline{X}_n - B \underline{X}_{n-1}) | \underline{X}_{n-1}] \text{ for every } \underline{X}_{n-1}.$$

Averaging over \underline{X}_{n-1} ,

$$\epsilon^2 = E \epsilon^2(\underline{X}_{n-1}) = E(\underline{X}_n - B \underline{X}_{n-1})^T (\underline{X}_n - B \underline{X}_{n-1})$$

must also be minimized by B . We will now prove that B is given by the extended orthogonality principle

$$E[(\underline{X}_n - B \underline{X}_{n-1}) \underline{X}_{n-1}^T] = 0. \quad (A1)$$

Indeed, consider any $B^* \neq B$. Then

$$\begin{aligned} \epsilon^{*2} &= E[(\underline{X}_n - B^* \underline{X}_{n-1})^T (\underline{X}_n - B^* \underline{X}_{n-1})] \\ &= E[(\underline{X}_n - B \underline{X}_{n-1}) + (B - B^*) \underline{X}_{n-1}]^T [(\underline{X}_n - B \underline{X}_{n-1}) + (B - B^*) \underline{X}_{n-1}] \\ &= \epsilon^2 + 2 E[(\underline{X}_n - B \underline{X}_{n-1})^T (B - B^*) \underline{X}_{n-1}] \\ &\quad + E[(B - B^*) \underline{X}_{n-1}]^T [(B - B^*) \underline{X}_{n-1}] \end{aligned}$$

But, from equation (A1), every component of $\underline{X}_n - B \underline{X}_{n-1}$ is orthogonal to every component of \underline{X}_{n-1} , so the second term above is zero. Furthermore the third term is always positive. Thus $\epsilon^{*2} \geq \epsilon^2$, and B minimizes ϵ^2 .

Solving equation (A1) for B ,

$$B = C_{n,n-1} C_{n-1,n-1}^T \quad (A2)$$

where

$$C_{n,k} = E \underline{X}_n \underline{X}_k^T = E \underline{X}(n\Delta) \underline{X}(k\Delta)^T = C(n\Delta, k\Delta).$$

The conditional covariance matrix is given by

$$\begin{aligned} K &= E[(\underline{X}_n - B \underline{X}_{n-1})(\underline{X}_n - B \underline{X}_{n-1})^T | \underline{X}_{n-1}] \\ &= E[(\underline{X}_n - B \underline{X}_{n-1})(\underline{X}_n - B \underline{X}_{n-1})^T] \\ &= E[(\underline{X}_n - B \underline{X}_{n-1}) \underline{X}_n^T] \end{aligned}$$

where the last two steps follow from condition (A1) and the further information that each component of $\underline{X}_n - B \underline{X}_{n-1}$ and each component of \underline{X}_{n-1} , since jointly normal, zero mean, and orthogonal, are independent. Thus

$$K = C_{n,n} - B C_{n,n-1}^T.$$

Using equation (A2),

$$K = C_{n,n} - C_{n,n-1} C_{n-1,n-1}^{-1} C_{n,n-1}^T. \quad (A3)$$

If the process $X(t)$ is stationary, equations (A2) and (A3) simplify to

$$B = C_{\Delta} C_o^{-1} \quad (A4)$$

and

$$K = C_o - C_{\Delta} C_o^{-1} C_{\Delta}^T. \quad (A5)$$

V.A. Wideband FM Generation

The classic methods for directly generating an FM signal are the reactance tube or varicap modulation of an LC oscillator. These methods suffer from both theoretical and practical limits so that to achieve less than 2 percent distortion one is normally restricted in both peak deviation, Δf , and maximum modulating frequency, f_m , to values below 10 percent of the carrier frequency. In many practical cases the restriction is to 1 percent or less of the carrier frequency.

The modulating frequency limitation can be overcome in these circuits only by moving to a higher center frequency. The peak deviation limitation can be overcome by repeated frequency multiplications (and if necessary by heterodyning down so that the desired final center frequency and deviation can be achieved at the same time). Such a frequency multiplication process also multiplies the inherent oscillator instabilities as well as the desired modulation.^{1, 2}

In the following paragraphs it is shown that an FM wave, with no theoretical limits on f_m or Δf (other than Δf shall not exceed the carrier frequency) may be generated by nonlinear waveshaping of an FM triangular wave. The defining equation for the triangular wave is first developed and then means of physically generating it are considered.

FM Triangular Wave

The most general form for a sinusoidal frequency modulated voltage wave, $v(t)$ is defined by

$$v(t) = A \sin \varphi(t) = A \sin \left[\int_0^t \omega_i(\tau) d\tau \right] \quad (1)$$

where $\varphi(t)$ is the instantaneous phase of the signal and $\omega_i(t) = \dot{\varphi}(t)$ is the instantaneous frequency. If $v(t)$ is to have a non-ambiguous form then $\omega_i(t) \geq 0$ is a necessary restriction. If $\omega_i(t)$ is written in the form $\omega_i(t) = \omega_o + \Delta\omega(t)$, this condition implies that the frequency deviation $\Delta\omega$ must be less than the carrier frequency ω_o .

Although when $v(t)$ is plotted versus time it is an aperiodic function, when it is plotted versus φ it takes the form of a periodic sinusoid of amplitude A . Similarly an FM triangular wave with the

same period as $v(t)$ and of amplitude B is defined as a periodic, symmetric triangular wave in the φ domain, i.e., when plotted versus φ . This waveform $T(t)$ in the φ domain and in the t domain is defined by the set of equations

$$\begin{aligned} T(t) &= \frac{2B\varphi(t)}{\pi} & : 0 < \varphi \leq \frac{\pi}{2} & : 0 < t \leq t_1 \\ &= \frac{2B}{\pi} [\pi - \varphi(t)] & : -\pi < \varphi \leq -\frac{\pi}{2} & : t_1 < t \leq t_2 \\ &= (-1)^n \frac{2B}{\pi} [\varphi - n\pi] & : (n - \frac{1}{2})\pi < \varphi \leq (n + \frac{1}{2})\pi & : t_n < t \leq t_{n+1} \end{aligned}$$

where t_n is the time at which $\varphi(t_n) = (n - \frac{1}{2})\pi$, and is shown in figure 1.

It should be noted that the FM triangular wave $T(t)$, as defined by equation 2, is not composed of a sequence of positive and negative ramps but rather is a waveform that is capable of varying considerably over the half-period $t_{n-1} < t < t_n$. Clearly such a wave has no theoretical maximum or minimum limit on its modulation frequency; i.e., the rate at which $\omega_1(t)$ is varied. Over every half-period, however, the slope must be monotonic since the instantaneous frequency $\omega_i(t)$ may never be negative without causing ambiguity in equation 1.

It turns out that it is possible both to generate $T(t)$ in a straightforward fashion and to convert $T(t)$ into $v(t)$ through the use of a non-memory, nonlinear network. This is quite apparent if we think of the nonlinear operation being performed in the φ domain.

For example, if $T(t)$ is passed through a network whose output v_o is related to its input v_i by the expression $v_o = A \sin(\frac{\pi v_i(t)}{2B})$ and if $v_i(t) = T(t)$ during the time interval $t_n \leq t \leq t_{n+1}$, then

$$\begin{aligned} v_o(t) &= A \sin [(-1)^n (-n\pi + \varphi(t))] \\ &= (-1)^n A \sin [-n\pi + \varphi(t)] = A \sin \varphi(t) \end{aligned}$$

Therefore, during this interval, $v_o(t) = v(t)$.

Since the network is memoryless the same result occurs for all time intervals; hence, $v_o(t) = v(t)$ for all values of n and for all time.

The nonlinear network described above may be synthesized in several different fashions.^{3,4} If hot carrier diodes are employed in a diode shaping network then the memoryless approximation will be valid

at least into the tens of megahertz region.

Theoretical Circuit Design

Figure 2 shows a block diagram for generating $T(t)$ from $\omega_i(t)$. The threshold sensor required in this diagram functions as follows. When the input voltage to the sensor reaches B it causes the switch to be thrown from contact a to contact b , and when the input reaches $-B$ the switch is thrown from contact b back to contact a . If at $t = 0$ the switch is on contact a , the output of the integrator is given by

$$v_i(t) = \frac{2B}{\pi} \int_0^t \omega_i(\tau) d\tau = \frac{2B}{\pi} \varphi(t)$$

When this output reaches B , which is at $t = t_1$ from equation 2, S is thrown and the output of the integrator then takes the form

$$\begin{aligned} v_i(t) &= \frac{2B}{\pi} \left[\int_0^{t_1} \omega_i(\tau) d\tau - \int_{t_1}^t \omega_i(\tau) d\tau \right] \\ &= \frac{2B}{\pi} \left[2 \int_0^{t_1} \omega_i(\tau) d\tau - \int_0^t \omega_i(\tau) d\tau \right] = \frac{2B}{\pi} [\pi - \varphi(t)] \end{aligned}$$

Similarly when $t = t_2$, S returns to a and $v_i(t) = -2\pi + \varphi(t)$. Comparison of $v_i(t)$ with equation 2 indicates that the block diagram of figure 2 does indeed generate $T(t)$ and $v(t)$.

A practical embodiment of the block diagram of figure 2 is shown in figure 3. This circuit uses a grounded capacitor as an integrator. The threshold sensor here is a Schmitt trigger designed with sufficient hysteresis (say 8-10 volts) to switch states to produce an output of V when its input reaches B and switch back to produce an output of $-V$ when its input reaches $-B$. When the Schmitt trigger output has a value of plus V , transmission through the controlled gate is possible, thereby supplying a current of $\frac{-2BC}{\pi} \omega_i(t)$ to the top of the integrating capacitor. When the Schmitt trigger output has a value of $-V$ no transmission is possible through the controlled gate; hence a current of $\frac{2BC}{\pi} \omega_i(t)$ is supplied to the top of the capacitor. Clearly the operation of this network is identical with the operation of the block

diagram of figure 2 . For any particular application, values of current, voltage, and capacity may be scaled as desired.

It is interesting to note that the network of figure 4 provides not only an FM triangular wave (which is shaped to produce the FM sinusoid) but also an FM square wave at the output of the Schmitt trigger. Consequently such a circuit provides a very useful test generator with no theoretical limitations on the instantaneous frequency. Circuit limitations seem to place about a 10 MHz upper bound on the instantaneous frequency with direct implementation of the network of figure 4.

Experimental Results

Several solid state realizations of the proposed type of generator have been constructed and tested. In a particular version of the circuit hot carrier diodes were used in the 12 diode sine wave shaping network to reduce diode frequency effects to a minimum. This version was designed to have a center frequency f_o variable in the range of 200 KHz to 2MHz; where f_o is the dc component of $\omega_1(t)$ divided by 2π . Over this range the harmonics of the carrier were suppressed by more than 46dB (to 0.5 percent or less of the fundamental).

Attempts to provide an accurate measurement of the generator's FM distortion have not been too successful since the distortion levels are very low⁵. One normally measures overall system distortions on the order of 0.2 percent for frequency deviations up to $0.75 f_o$. Since both the input sine wave modulating generator and the broadband FM pulse count receiver are known to have distortions of this order of magnitude we can only speculate as to the exact distortion produced by the generator itself.

Figures 4 and 5 show outside and internal views of a version of this generator that is now in use in our laboratories.

Footnotes

- (1) M. Schwartz, Information Transmission, Modulation and Noise, McGraw Hill, 1959, Chapter 3.
- (2) M. Ogi, and T. Sekizawa, "Solid State Broadband FM Modulator and Demodulator". FUJITSU Scientific and Technical Journal, Vol. 2, No. 2, September 1966, p. 27-47.
- (3) L. Strauss, "Wave Generation and Shaping", McGraw Hill, 1960, Section 208 and Problem 2-22.
- (4) G. Klein, "Accurate Triangle-Sine Converter", Digest of Technical Papers, 1967 Int. Solid State Circuits Conference, Philadelphia, Feb. 1967, p. 120-121.
- (5) D.M. Petrela, "A Wideband FM Generator", Master of Science Report Project, Electrical Engineering Department, Polytechnic Institute of Brooklyn, 1966. Circuit diagrams and detailed measurements for a circuit of the type described.

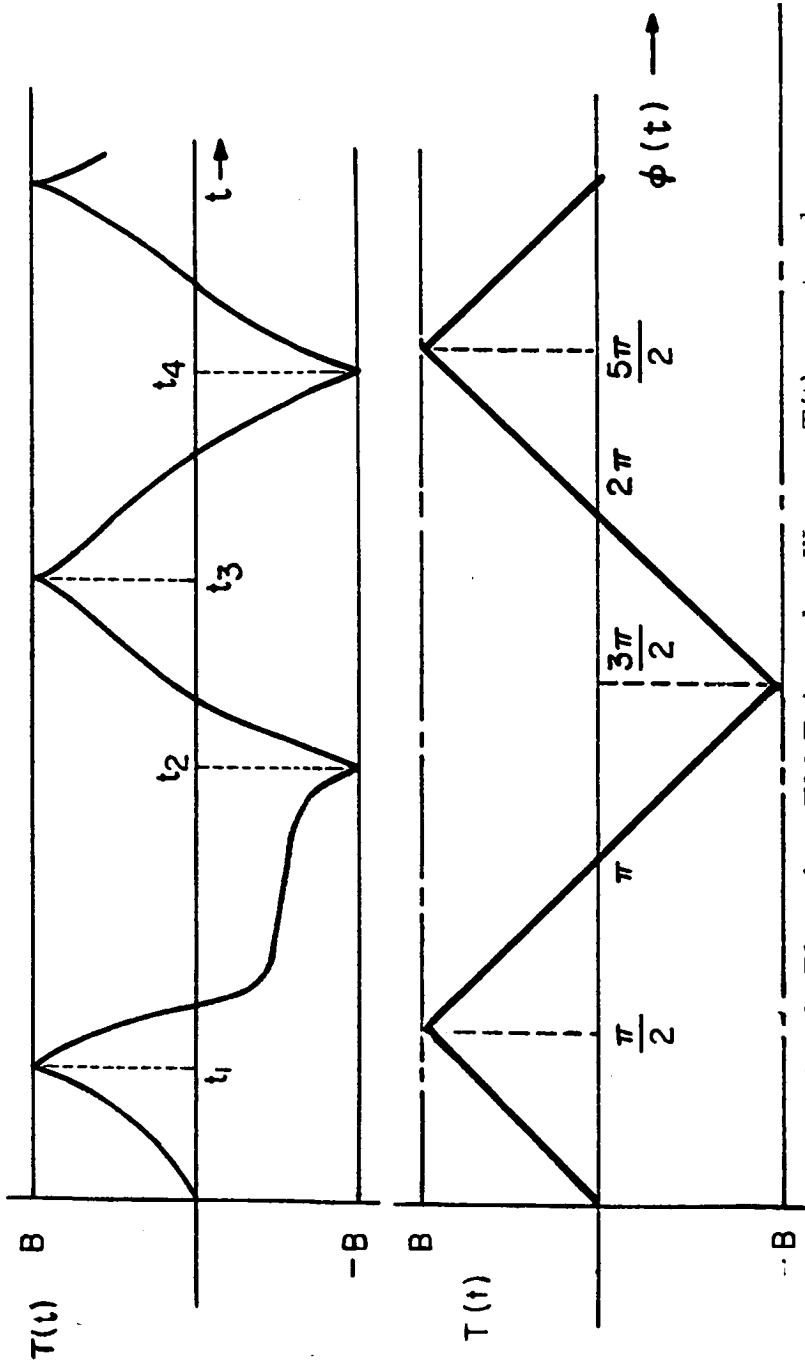


Fig. 1 Plot of an FM Triangular Wave, $T(t)$, vs. t and ϕ

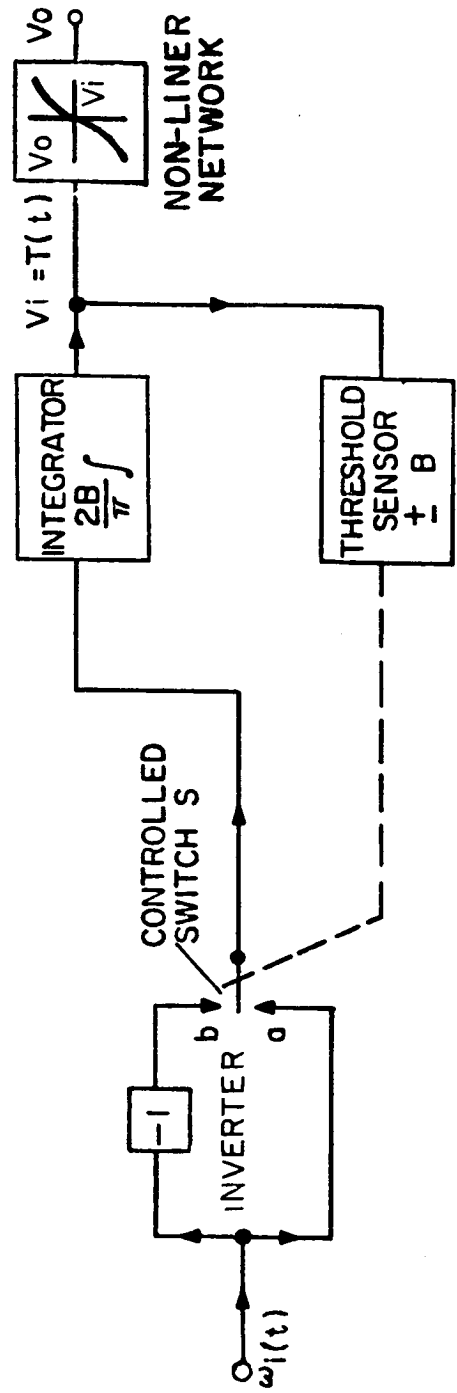


Fig. 2 Block Diagram for Generating $T(t)$

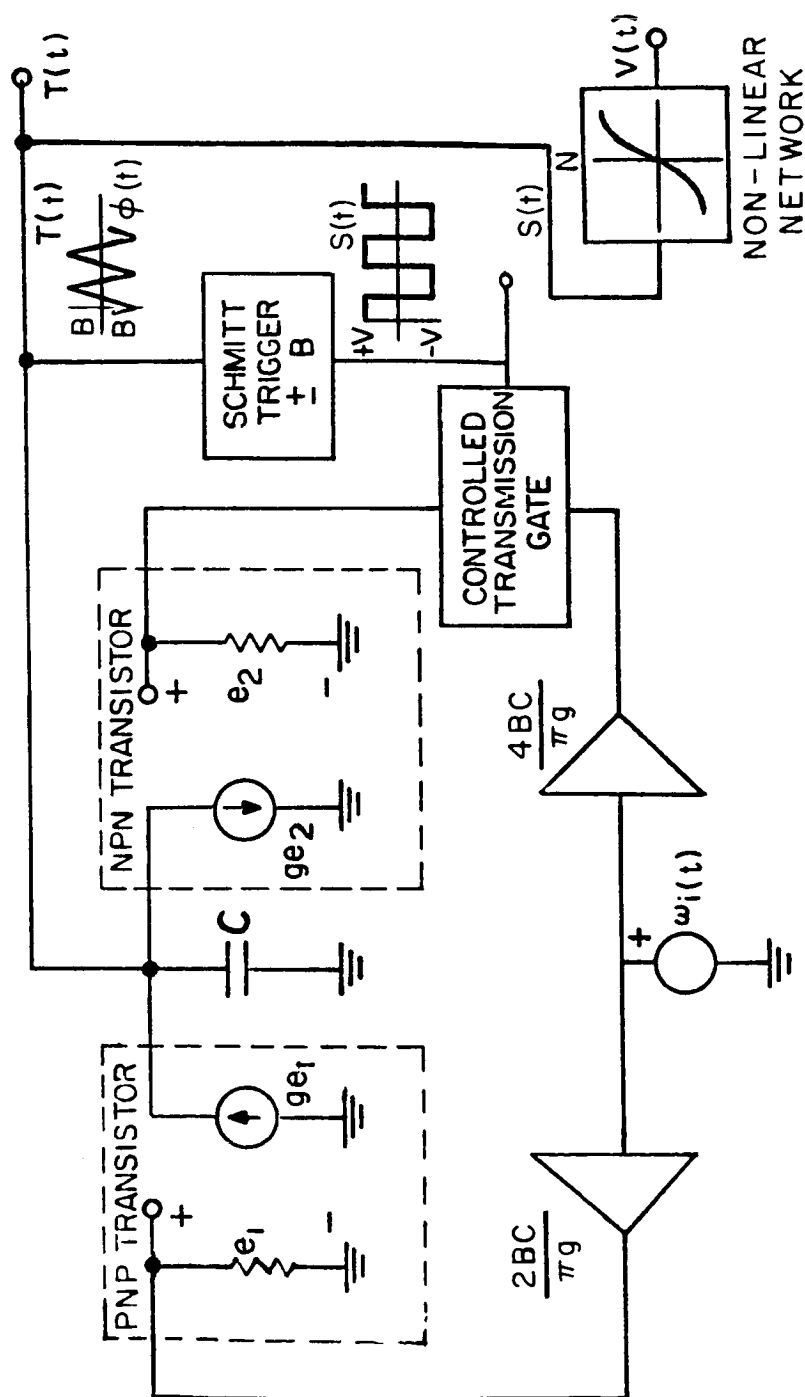


Fig. 3 Practical Network for Generating $T(t)$ from $\omega_i(t)$

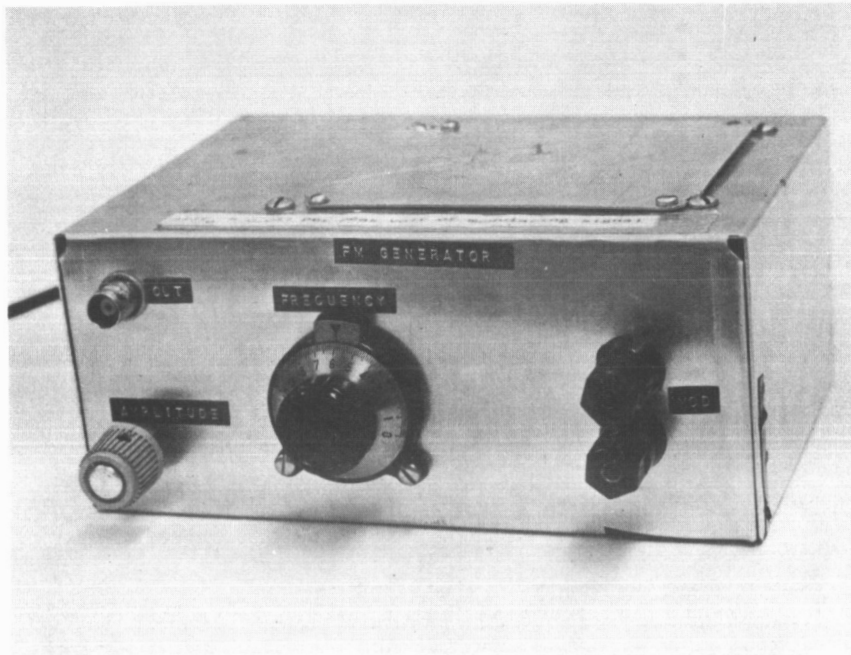


Fig. 4 Solid State, Wideband, Low Distortion
FM Generator

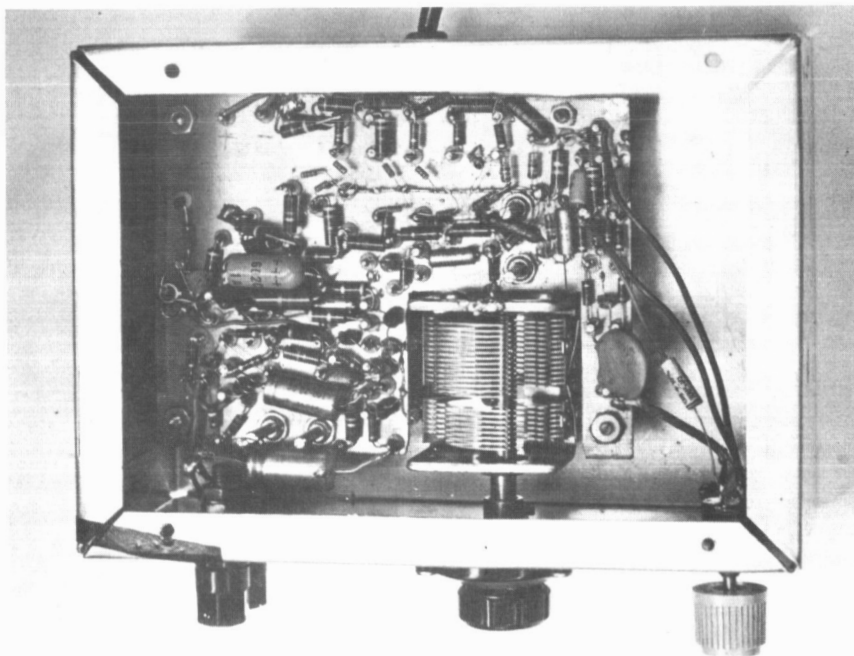


Fig. 5 Internal view of FM Generator

V.B. SNR For Bandpass Limiters*

Introduction

The bandpass limiter is a nonlinear device which ideally eliminates any amplitude modulation from an input waveform. Bandpass limiters are used extensively in FM detection to remove amplitude fluctuations from the FM signal before discrimination. In addition, the triggering of a Schmitt trigger or Multivibrator with a noisy signal is a second widely used application.

When limiters are used to provide limiting of noisy signals, the resulting signal is frequency modulated. The frequency jitter present in this signal has been determined by Rice^{1, 2} and others.

It is often important to decompose this "FM" signal into its carrier and sidebands. The carrier (if the signal is unmodulated) represents the "true" output signal while the power in the sidebands represent the output noise power. An exact expression for the output signal to noise ratio of an ideal limiter has been obtained by Davenport⁽³⁾ and for a smooth limiter by Galejs⁽⁴⁾. However, each of these derivations are extremely complicated and the resulting expressions are given in series expansions involving hypergeometric functions.

The technique given below leads to an approximate result, easily derived, and having a closed form solution. In addition, the technique used has wide applicability to many types of FM problems.^(5, 6)

Review of the Theory

A block diagram of the system considered is shown in Fig. 1. The input voltage, $v(t)$ consists of a sinusoidal signal embedded in white gaussian noise:

$$v(t) = A \cos \omega_o t + n_w(t) \quad (1)$$

This voltage is filtered as shown. The voltage, $V_i(t)$, at the input to the limiter is therefore:

$$\begin{aligned} V_i(t) &= A \cos(\omega_o t) + x(t) \cos \omega_o t - y(t) \sin \omega_o t \\ &= A \cos \omega_o t + r(t) \cos(\omega_o t + \theta(t)) \end{aligned} \quad (2)$$

* This report is to be published in the IEEE Transactions on Aerospace and Electronic Systems in January 1968.

where it has been assumed that the filter does not distort the signal in any way while shaping the noise, and that $x(t)$, and $y(t)$ are independent, normal processes each with mean zero, and variance, N_i .

Combining signal and noise in Eq. 2 yields:

$$V_i(t) = R(t) \cos(\omega_o t + \varphi_n(t)) \quad (3a)$$

where

$$R = \left[(A + x)^2 + y^2 \right]^{1/2} \quad (3b)$$

and

$$\varphi_n = -\tan^{-1} \frac{y}{A+x} \quad (3c)$$

Rice's technique⁽²⁾ represents the phase noise, $\varphi_n(t)$, by two terms. The first term is obtained by assuming that at high signal to noise ratios the noise components are usually much smaller than the carrier amplitude, A . When this is true, Eq. 3c can be written as

$$\varphi_g = -\frac{y}{a} \quad (4)$$

This component of phase noise is normally distributed with mean zero and variance, $\frac{1}{2} \left(\frac{N_i}{S_i} \right)$, (where $S_i = A^2/2$).

The second term is best described with the aid of the phasor representation of Eq. 2, shown in Fig. 2. When $r(t)$ exceeds, A , it becomes possible for $\varphi_n(t)$ to rotate 2π radians. The conditions required for this rotation are:

1. $r(t) > A$
2. $\theta(t)$ going through π
3. $\dot{\theta}(t)$ greater than zero

If $\theta(t)$ were less than zero, the resulting clockwise rotation would cause a negative 2π jump in phase. The rate of occurrence of these jumps were assumed by Rice to be governed by a Poisson distribution. This conjecture was verified experimentally by Ringdahl and Schilling⁽⁷⁾ for SNR above 4db.

Rice's technique completely ignores the "in between" noise conditions. However, the results obtained using this technique to determine the output SNR of a Discriminator closely check Rice's earlier work⁽²⁾. In addition, this technique also was used successfully to determine error rates obtained using an FM Discriminator Detector⁽⁶⁾. In both cases it was found that the technique is applicable for SNR exceeding 3db.

To express this second term, to be called $\varphi_s(t)$, analytically, the following assumption is made:

Each $\pm 2\pi$ jump in $\varphi_s(t)$ can be approximated by a ramp of width τ_s and height $\pm 2\pi$, as shown in Fig. 3a. An equivalent representation for $\varphi_s(t)$ is possible by requiring the ramps to be modulo 2π . Fig. 3b shows the resulting process as a train of triangular-shaped pulses. Since the pulses are (assumed to be) poisson distributed in time, the process can be represented as a "shot noise" effect:

$$\varphi_s(t) = \sum_{i=-\infty}^{\infty} b_i h(t-t_i) \quad (5a)$$

where

$$b = \pm 1 \quad (5b)$$

$$P(b=+1) = P(b=-1) = \frac{1}{2} \quad (5c)$$

and

$$h(t) = \begin{cases} \frac{2\pi t}{\tau_s} = \omega_s t & 0 \leq t \leq \tau_s \\ 0 & \text{elsewhere} \end{cases} \quad (5d)$$

It is easily shown that the expected number of pulses occurring per second, λ , is⁽²⁾

$$\lambda = B_{\text{rms}} \operatorname{erfc} \sqrt{\frac{S_i}{N_i}} \quad (6)$$

where B_{rms} is the rms bandwidth of the input filter.

Since the pulses have duration of τ_s seconds and the noise is shaped by a filter with a bandwidth B_{rms} one finds experimentally that the average value of τ_s is inversely proportional to B_{rms} .

It was found by comparing the exact and approximate results that

$$\frac{1}{B_{\text{rms}}} \leq \tau_s \simeq \frac{2}{B_{\text{rms}}} \quad (7)$$

Some New Results

The input signal is passed through the ideal limiter and an output filter which removes higher harmonics, (as $2f_o$, $3f_o \dots$). The output of this filter is then,

$$V_o(t) = \cos(\omega_o t + \varphi_g(t) + \varphi_s(t)) \quad (8)$$

where the output amplitude is assumed to be unity and the output filter is wide enough not to cause any distortion of $V_o(t)$.

The procedure now employed consists of calculating the correlation function for $R_{V_o}(\tau)$ and its Spectral Density. The Signal and Noise powers are then separated and the output SNR determined. Thus

$$\begin{aligned} R_{V_o V_o}(t, \tau) &= E(v_o(t+\tau) v_o(t)) = \\ &= \frac{1}{2} R_e \left\{ e^{j\omega_o \tau} E \left[e^{j(\varphi_g(t+\tau) - \varphi_g(t))} \right] E \left[e^{j(\varphi_s(t+\tau) - \varphi_s(t))} \right] \right\}^{(9)} \end{aligned}$$

The ensemble averages in Eq. 9 are recognized as the joint characteristic functions of the Gaussian and shot noise:

$$C_g(u, v) = E \left\{ e^{j[u\varphi_g(t-\tau) + v\varphi_g(t)]} \right\} \quad (10a)$$

$$C_s(u, v) = E \left\{ e^{j[u\varphi_s(t+\tau) + v\varphi_s(t)]} \right\} \quad (10b)$$

Thus,

$$R_{V_o V_o}(t, \tau) = \frac{1}{2} R_e \left\{ e^{j\omega_o \tau} C_g(1, -1) C_s(1, -1) \right\} \quad (11)$$

where

$$C_g(u, v) = e^{-\frac{1}{2} [u^2 \sigma_{\varphi_g}^2 + 2uv R_{\varphi_g}(\tau) + v^2 \sigma_{\varphi_g}^2]} \quad (12)$$

Hence

$$C_g(1, -1) = e^{-\sigma_{\varphi_g}^2 (1 - \rho(\tau))} \quad (13a)$$

where

$$\sigma_{\varphi_g}^2 = \frac{N_i}{2S_i} \quad (13b)$$

and

$$\rho(\tau) = \frac{1}{2} \frac{R_{\varphi_g}(\tau)}{\sigma_{\varphi_g}^2} \quad (13c)$$

The joint characteristic function of Poisson distributed shot noises is given by Gilbert + Pollak⁽⁸⁾:

$$C_s(u, v) = \exp \left[\lambda \int_{-\infty}^{\infty} \left(E \left\{ e^{jb [uh(\beta+\tau) + vh(\beta)]} \right\} - 1 \right) d\beta \right] \quad (14)$$

Letting $u = 1$ and $v = -1$ yields

$$C_s(1, -1) = \exp \left[\lambda \int_{-\infty}^{\infty} \left(E \left\{ e^{jb [h(\beta+\tau) - h(\beta)]} \right\} - 1 \right) d\beta \right] \quad (15)$$

However,

$$E \left\{ e^{jb [h(\beta+\tau) - h(\beta)]} \right\} = \int_{-\infty}^{\infty} e^{jb [h(\beta+\tau) - h(\beta)]} f_b(b) db \quad (16)$$

where

$$f_b(b) = \frac{1}{2} [\delta(b+1) + \delta(b-1)]$$

Evaluating eq. 16 yields

$$C_s(1, -1) = \exp \left\{ \lambda \int_{-\infty}^{\infty} (\cos [h(\beta+\tau) - h(\beta)] - 1) d\beta \right\} \quad (17)$$

Hence,

$$C_s(1, -1) \begin{cases} \exp \lambda \left(\frac{2}{\omega_s} \sin \omega_s |\tau| + (\tau_s - |\tau|) \cos \omega_s \tau - (\tau_s + |\tau|) \right) & |\tau| \leq \tau_s \\ \exp [-2\lambda \tau_s] & |\tau| \geq \tau_s \end{cases} \quad (18)$$

Substituting Eqs. 13a and 18 into Eq. 11 and expanding one finds that:

$$\begin{aligned} R_{v_o v_o}(\tau) = \frac{1}{2} e^{-[\sigma_{\varphi}^2 + \lambda \tau_s]} \cos \omega_o \tau & \left\{ 1 + \sum_{k=1}^{\infty} \frac{R_{\varphi}^k(\tau)}{k!} + \right. \\ & + \sum_{m=1}^{\infty} \frac{\lambda^m \left[\frac{2}{\omega_s} \sin \omega_s |\tau| + (\tau_s - |\tau|) \cos \omega_s \tau - |\tau| \right]^m}{m!} + \\ & \left. + \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{R_{\varphi}^k(\tau)}{k!} \cdot \frac{\lambda^m \left[\frac{2}{\omega_s} \sin \omega_s |\tau| + (\tau_s - |\tau|) \cos \omega_s \tau - |\tau| \right]^m}{m!} \right\} \quad (19) \\ & \left. \right\} |\tau| < \tau_s \end{aligned}$$

The first term within the brackets represents the signal term, the second term (\sum_k) is the Gaussian noise term centered at ω_o , the third term (\sum_m) is the shot noise term centered at ω_o , and the fourth term $(\sum_k \sum_m)$ is the interaction of shot noise and Gaussian noise.

Since we are only interested in the output signal and noise power we need not take the spectral density of eq. 19. However, we see that the spectral density is easily obtained. The output signal power is

$$S_o = \frac{1}{2} e^{-(\sigma_{\phi_g}^2 + \lambda \tau_s)} = \frac{1}{2} e^{-\left[2 \frac{1}{S_i/N_i} + \tau_s B_{rms} \operatorname{erfc} \sqrt{\frac{S_i}{N_i}} \right]} \quad (20)$$

Thus, the signal power is $\frac{1}{2}$ when there is no noise ($\sigma_{\phi_g}^2 = 0$ and $\lambda = 0$).

The output noise power is easily calculated from Eq. 19 by letting $\tau = 0$ in the noise terms:

$$\begin{aligned} N_o &= \frac{1}{2} e^{-[\sigma_{\phi_g}^2 + \lambda \tau_s]} \left\{ \sum_{k=1}^{\infty} \frac{\sigma_{\phi_g}^{2k}}{k!} + \sum_{m=1}^{\infty} \frac{\lambda^m \tau_s^m}{m!} + \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sigma_{\phi_g}^{2k}}{k!} \cdot \frac{\lambda^m \tau_s^m}{m!} \right\} \\ &= \frac{1}{2} e^{-[\sigma_{\phi_g}^2 + \lambda \tau_s]} \left\{ (e^{\sigma_{\phi_g}^2} - 1) + (e^{\lambda \tau_s} - 1) + (e^{\sigma_{\phi_g}^2} - 1)(e^{\lambda \tau_s} - 1) \right\} \end{aligned} \quad (21)$$

Simplifying:

$$N_o = \frac{1}{2} \left[1 - e^{-\left\{ \frac{1}{2S_i/N_i} + B_{rms} \tau_s \operatorname{erfc} \sqrt{\frac{S_i}{N_i}} \right\}} \right] \quad (22)$$

If we add Eqs. 20 and 22, we find:

$$S_o + N_o = \frac{1}{2} \quad (23)$$

This result is expected from Eq. 8, where it is seen that the total power contained in V_o is $\frac{1}{2}$.

The output SNR is

$$\frac{S_o}{N_o} = \frac{1}{e^{\left[\frac{1}{2S_i/N_i} + \tau_s B_{rms} \operatorname{erfc} \sqrt{\frac{S_i}{N_i}} \right]} - 1} \quad (24)$$

It is seen from Eq. 24 that the output SNR decreases as the input noise increases. Referring to Eq. 8, it is seen that when no noise is present the total power in the sidebands increases and hence the signal power decreases. The sideband power is, in this case, the output noise power. Notice the closed form solution of Eq. 24.

The signal suppression effect given by eq. 20 is compared with the exact solution of Davenport in Fig. 4 for the cases of $B_{\text{rms}} \tau_s$ equal to 1, 2, and 3. Note that $B_{\text{rms}} \tau_s$ equal to 1 or 2 represents the "best" match to the exact curve. Figure 5 shows a comparison of output and input SNR for $B_{\text{rms}} \tau_s = 1, 2$, and 3. Note that the approximation begins to break down for input SNR below 3db. However, Rice's approximation is no longer valid below 3db.

Conclusions

Using Rice's approach to FM, a closed form solution to the ideal limiter problem is obtained. It is shown that the results compare quite well with the exact results for input SNR above 3 db.

References

1. Rice, S.O., "Statistical Properties of a Sine-Wave Plus Random Noise, " BSTJ, January 1948.
2. Rice, S.O., "Noise in Fm Receivers", Chapt. 25 of Time Series Analysis, edited by Rosenblatt, Wiley 1963.
3. Davenport, W.B., "Signal to Noise Ratios in Band-Pass Limiters", J. Appl. Physics, June 1953.
4. Galejs, J., "Signal to Noise Ratios in Smooth Limiters", IRE Trans. on Information Theory, June 1959.
5. Schilling, D.L., Nelson, E., and Clarke, K.K., "Response of a Fading FM Signal", IEEE Trans. on Comm. Tech, April 1967.
6. Schilling, D.L., Hoffman, E., and Nelson, E., "Error Rates in FSK Signals Using an FM Discriminator Detector", IEEE Trans. on Comm. Tech., August 1967.
7. Ringdahl, I., and Schilling, D.L., "On the Distribution of the Spikes Seen at the Output of an FM Discriminator below Threshold", Proceedings of IEEE, October 1964.
8. Gilbert, E.N., and Pollack, H.O., "Amplitude Distributions of Shot Noise", BSTJ, March 1960.

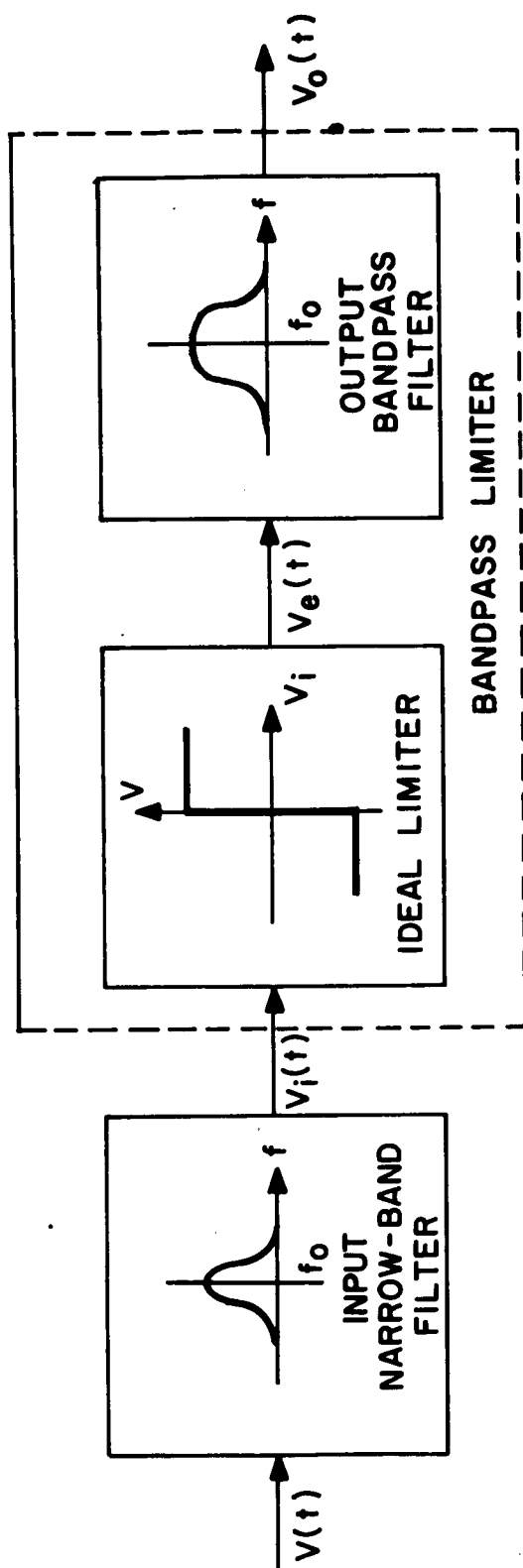


Figure 1

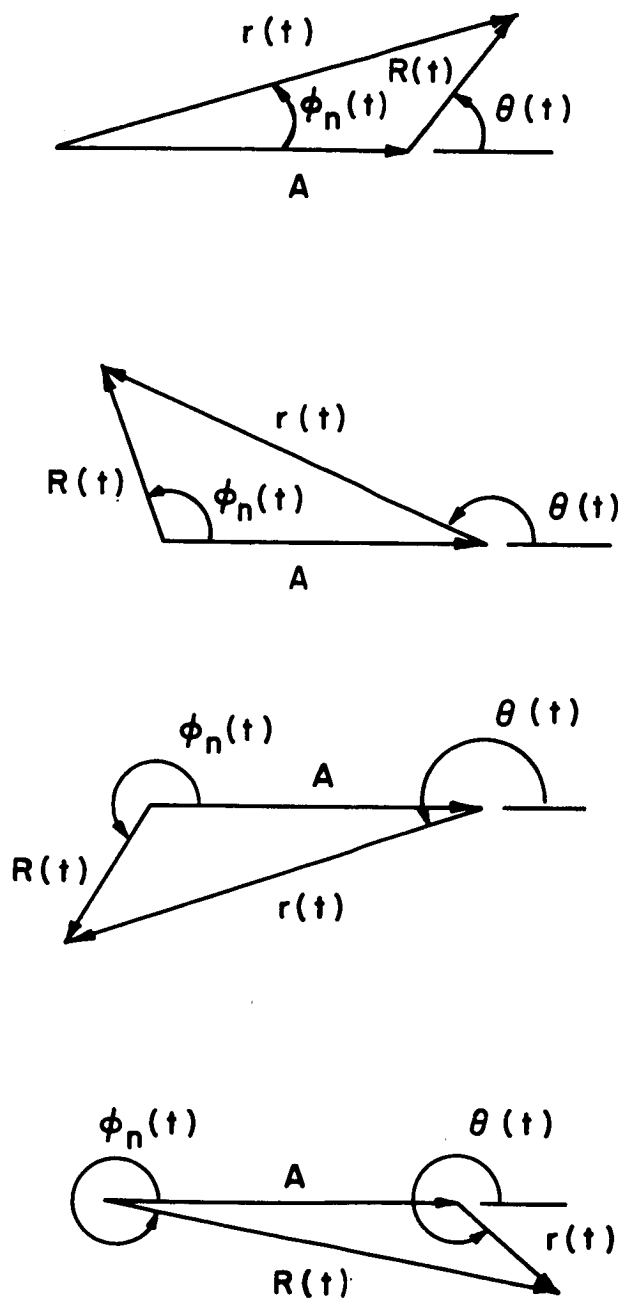
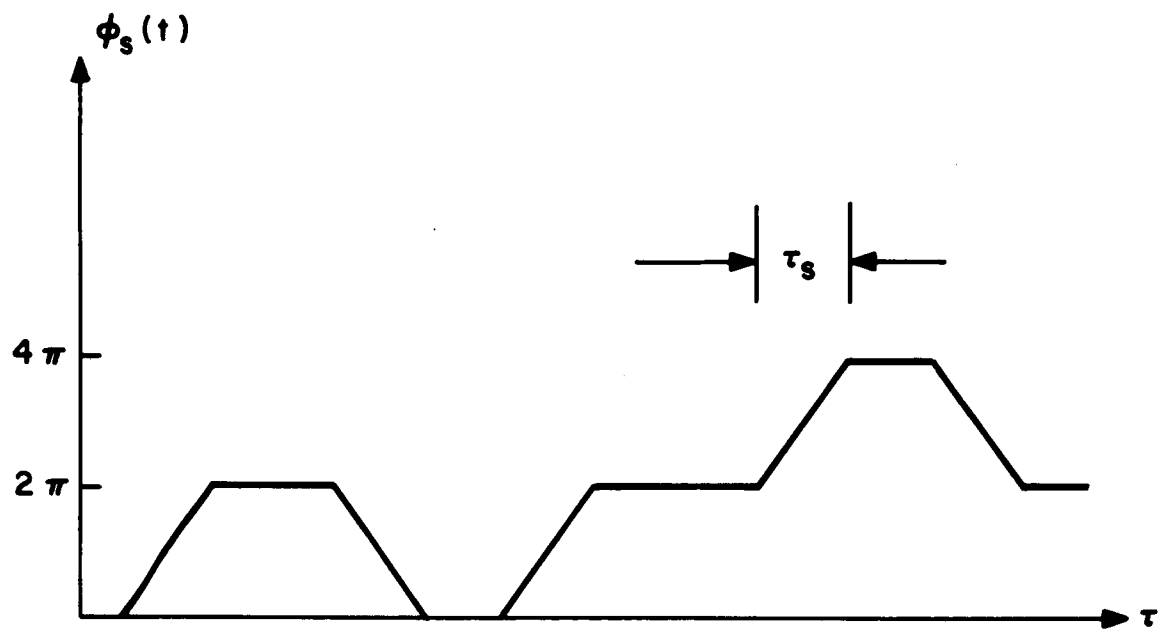
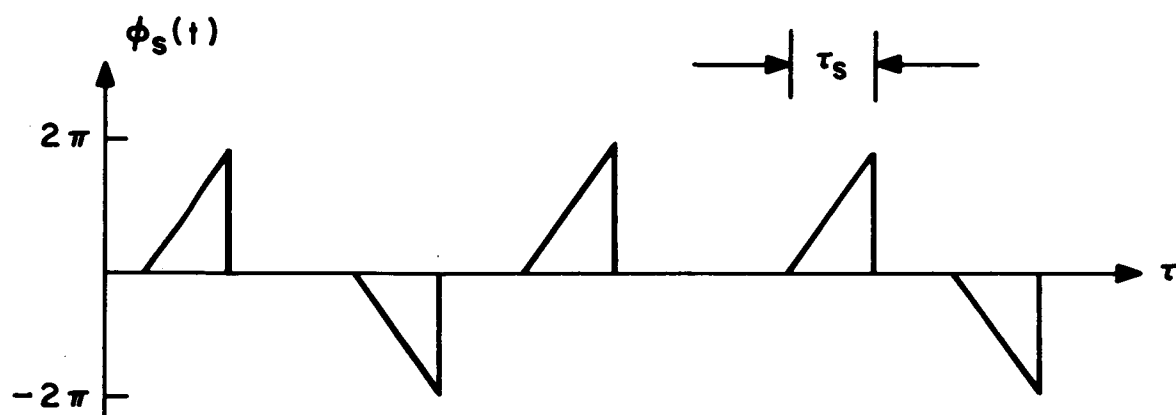


Figure 2



(a) Ramp Function Model



(b) Shot - Noise Model

Figure 3

BANDPASS LIMITER OUTPUT SIGNAL POWER CHARACTERISTICS

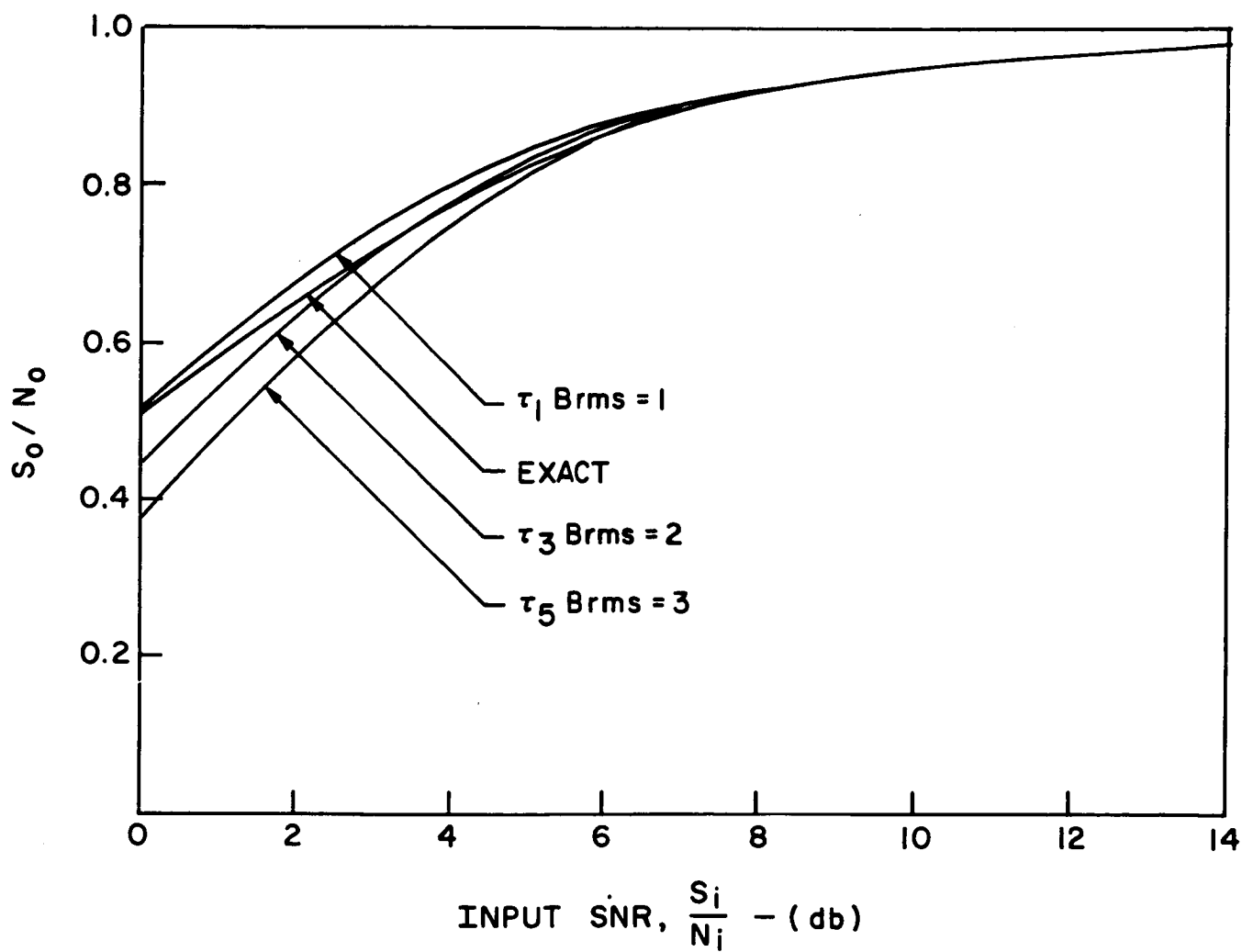


Figure 4

BANDPASS LIMITER SIGNAL - TO - NOISE RATIO CHARACTERISTICS

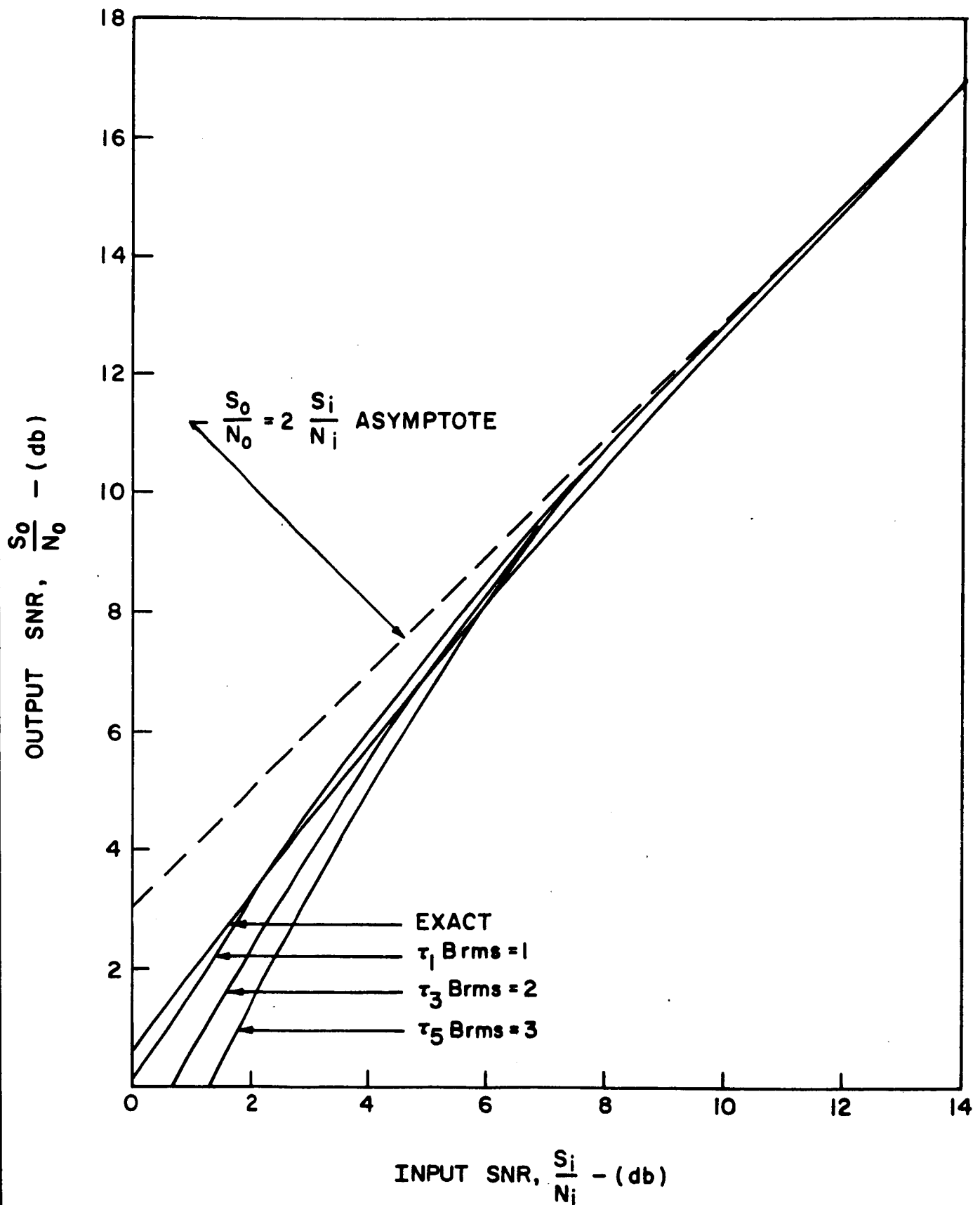


Figure 5

V.C. A Statistical Approach to Interference and Multipath in FM.

Although frequency-modulation communications system contain interference-suppression capability greater than that of amplitude-modulation systems of comparable power, FM systems are not, unfortunately, completely immune to such interference. Interfering noise, unrelated transmitters, or echos of the desired signal can produce interference in the demodulated information. Such factors can result in degradation of system performance.

While interference in AM systems results in such effects as beat notes or simultaneous reception of two signals, FM interference has somewhat different characteristics. Severe interference in FM occurs when the desired signal and the interfering signal combine to create sudden, rapid changes in the phase of the signal fed to the receiver, usually accompanied by a sudden decrease in amplitude of the carrier signal. At the receiver output, this results in a characteristic amplitude surge of "spike". As the interfering signal becomes stronger, the spikes become more severe. Should the interference become even stronger than the desired signal, the information of the desired signal is lost, and the information received will be that modulated on the stronger signal. The weaker signal causes unintelligible spike interference in the demodulated information of the stronger signal. The predominacy of the information of the stronger signal is known as the "capture effect".

When a system is plagued with such interfering signals, it is desirable to characterize mathematically the interference and to make some attempt to predict the severity of the effect on the performance of the system.

This opens the way to attempts to decrease or eliminate the effects of the interference. An obvious first step is to write the equation of the waveform of the interference. This may be simple, but the expressions may be arduous. Furthermore, the interference may depend upon factors which cannot be predicted exactly in advance. An example might be the strength or phase of an interfering signal. Frequently a statistical approach is desirable in the treatment of such systems.

To illustrate the character of the "spike" interference encountered in FM and the application of statistical techniques, we consider first a very simple case. Suppose we are not transmitting information, but only a sinusoidal carrier wave. We suppose our demodulator or frequency discriminator to be "ideal", i. e., its output equals to instantaneous frequency of the input. We will measure this output in radians per second. In the case of a single sinusoidal input, then, the receiver output is just the frequency of the input wave.

Now, suppose another transmitter broadcasts another unmodulated sinusoid of frequency different from the frequency of the first, but not greatly different. The receiver picks up the sum of the two signals, $\sin(\omega, t) + \rho \sin(\omega', t)$. The second signal creates interference, and, instead of the constant output desired, a periodic train of spikes results. The period is equal to the period of the difference frequency of the two inputs. The amplitude of the spikes depends upon the signal ratio, ρ ; figure 1(a) shows the receiver output for signal ratios of 0.5, 0.7, and 0.8. The output for two cases of $\rho > 1$ is also shown; in this instance, the "echo" has become the stronger signal.

Valuable information about this interference process can be gained by time averages of the demodulated signal, $\omega_r(t)$. Some important averages are shown below.

$$\omega_r(t) = \omega_1, \rho < 1$$

$$\omega_r(t) = \omega_2, \rho > 1$$

$$\overline{(\omega_r(t) - \omega_r(t))^2} = \frac{(\omega_2 - \omega_1)^2}{2} \frac{\rho^2}{1 - \rho^2}, \rho < 1$$

$$\overline{(\omega_r(t) - \omega_r(t))^2} = \frac{(\omega_2 - \omega_1)^2}{2} \frac{1}{\rho^2 - 1}, \rho > 1$$

Note that the average output is always equal to the frequency of the stronger signal. Thus, a low-pass filter of sufficiently narrow bandwidth will always filter out the desired (constant) output, provided the receiver is not captured by a stronger signal. In addition, the power of the interference is unbounded in the region about $\rho = 1$. For $\rho = 1$, in fact, step discontinuities in the phase of the total signal occur, and an ideal receiver produces impulses at its output. In any real system, of course, the receiver will somehow limit the amplitude of the output.

One way to apply statistical methods to this two-carrier interference is to consider the signal amplitude ratio, ρ , as a random variable. It is likely that this parameter cannot be predicted in advance. We could assume a Raleigh density (figure 1(b)) for ρ , as might occur in atmospheric reflection. It is then possible to calculate a density function for the resulting ensemble of waveforms; this density function will be a time function. Figure 1(c) shows how the density function appears at various points in the cycle of the spike train (fig. 1(a)). Thus, we can observe the distribution of the tips of the spikes (180° point in cycle), and also the density at other points in the cycle.

The foregoing example is not of practical interest: no information was transmitted. The waveform (fig. 1(a)), however, is of general importance in situations where just two interfering signals occur. The argument of the function will, in general, depend not linearly on time but as a more complicated time function of the modulation: the instantaneous phase difference of the two interfering signals. One application is multipath transmission, in which the interfering signal is a time-delayed echo of the desired signal.

Suppose the received signal is the sum of the principal transmitted signal

$$e_1(t) = \cos [\omega_c t + \underline{\Phi}(t)]$$

and a single echo delayed in time by Δt :

$$e_2(t) = \rho \cos [\omega_c (t - \Delta t) + \underline{\Phi}(t - \Delta t) + \phi]$$

If the echo is less strong than the principal transmitted signal, it can be shown that the discriminator will produce an output

$$\omega_r(t) = \omega_c + \frac{d}{dt} \underline{\Phi}(t) + \frac{d}{dt} \left[\tan^{-1} \frac{\rho \sin \psi}{1 + \rho \cos \psi} \right], \text{ where } \psi = \underline{\Phi}(t - \Delta t) - \underline{\Phi}(t) + \phi - \omega_c \Delta t.$$

In ω_r , the first term is a constant (which can be eliminated by use of a balanced discriminator), the second term is the desired signal, and the third term, to be called ω_i , the unwanted interference. Simplifying

$$\text{we get: } \omega_i = \frac{d}{dt} \tan^{-1} \frac{\rho \sin \psi}{1 + \rho \cos \psi} = \rho \frac{\rho + \cos \psi}{1 + \rho^2 + 2\rho \cos \psi} \cdot \frac{2\psi}{2t}.$$

As has been mentioned, the terms dependent upon ρ are identical to those occurring in the simple case of two unmodulated carriers, with the exception that the argument is now ψ . In addition, the interference is modulated by the derivative of the phase difference.

Calculation of the phase difference ψ can be difficult. If the time delay is small compared to the modulation, however, the phase difference can be approximated by the derivative of the modulation phase, which is just the original modulation information. Estimation of the spike density in the interference is then easy; the spike amplitude, however, is modulated by the derivative of the information and thus is more difficult to analyze. Another case that affords simple calculation of exact results is that in which the modulation is monochromatic: $\bar{\phi}(t) = \beta \sin \rho t$. In this case, the phase difference is simply a sinusoid, since it is the difference of two sinusoids of the same frequency:

$$\psi = \phi - \omega_c \Delta t - (2\beta \sin \rho \frac{\Delta t}{2}) \cos \rho(t - \frac{\Delta t}{2})$$

A graphical approach to the evaluation of these functions is helpful. In fig. 2(a), the function $\frac{\rho + \cos \psi}{1 + \rho^2 + 2\rho \cos \psi}$ is plotted on a horizontal scale with its argument, ψ , on the vertical scale. To the right of this plot is the actual modulation phase difference, ψ , as a function of time. For the specific instance in question, this is a sinusoid; it is not in general. By reading across between the two graphs, one can determine the location of the spikes; the resulting spikes are then plotted below as a time function, fig. 2(b). Finally, the spikes are modulated by the derivative of the phase difference, fig. 2(c), and added to the information signal to produce the actual receiver output, fig. 2(d).

This graphical approach gives insight into the spike patterns under arbitrary modulation conditions. If the phase difference can be approximated by the derivative of the instantaneous phase, data on spike frequency can be obtained for arbitrary, even random, modulation.

One problem in the characterization of interference is that the net average phase difference between the two carriers, $\theta = \phi - \omega_c \Delta t$, is not usually determined. A small change in time delay results in a significant change in the spike pattern. In fig. 2(e) is plotted the new interference that results if θ is changed by 180° from the case of fig. 2(c). Note that, while the spikes have changed location, the envelope has not.

One way to deal with this indeterminacy is to consider the average phase difference to be a random variable, evenly distributed between 0 and 2π . The resulting interference then becomes a nonstationary random process. The mean and variance of the process can be obtained by evaluation of the appropriate integrals, a straightforward but lengthy process. The results are shown below for $\rho < 1$. For $\rho > 1$, the significance of the "echo" and the principal signal are reversed.

$$E(\omega_i(t, \theta)) = \frac{1}{2\pi} \int_0^{2\pi} \omega_i(t, \theta) d\theta = 0$$

$$E([\omega_i(t, \theta)]^2) = \frac{1}{2\pi} \int_0^{2\pi} [\omega_i(t, \theta)]^2 d\theta = \frac{1}{2} (\Delta\omega \sin \rho \frac{\Delta t}{2})^2 \frac{\rho^2}{1-\rho^2} \sin^2 \rho(t - \frac{\Delta t}{2})$$

The standard deviation, or square root of the variance, is superimposed on fig. 2(e) as a broken line; the solid line is, of course, one member of the ensemble being averaged.

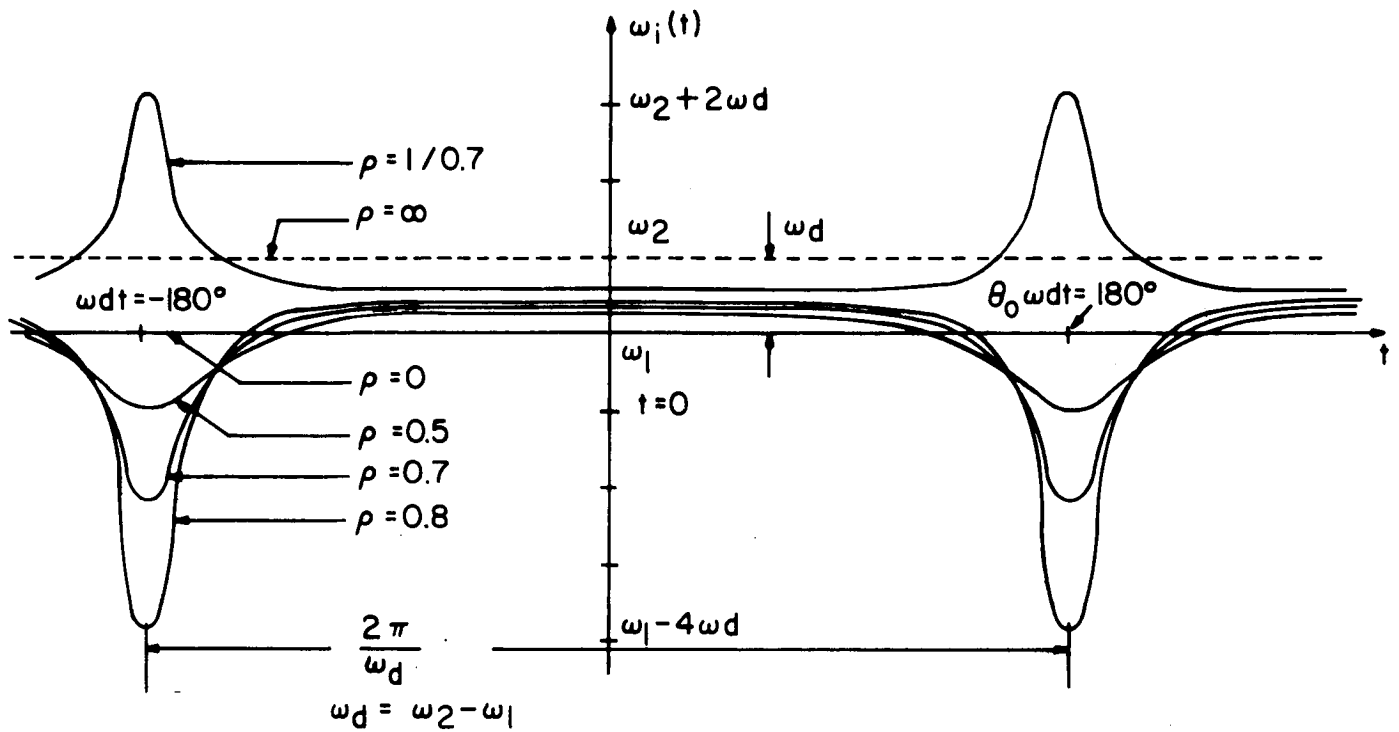
The variance may further be averaged over time to obtain an average mean-square power:

$$\frac{\rho}{2\pi} \int_0^{2\pi/\rho} \frac{1}{2\pi} \omega_i^2 d\theta dt = \frac{1}{2\pi} \int_0^{2\pi} \overline{[\omega_i(\theta, t)]^2} d\theta =$$

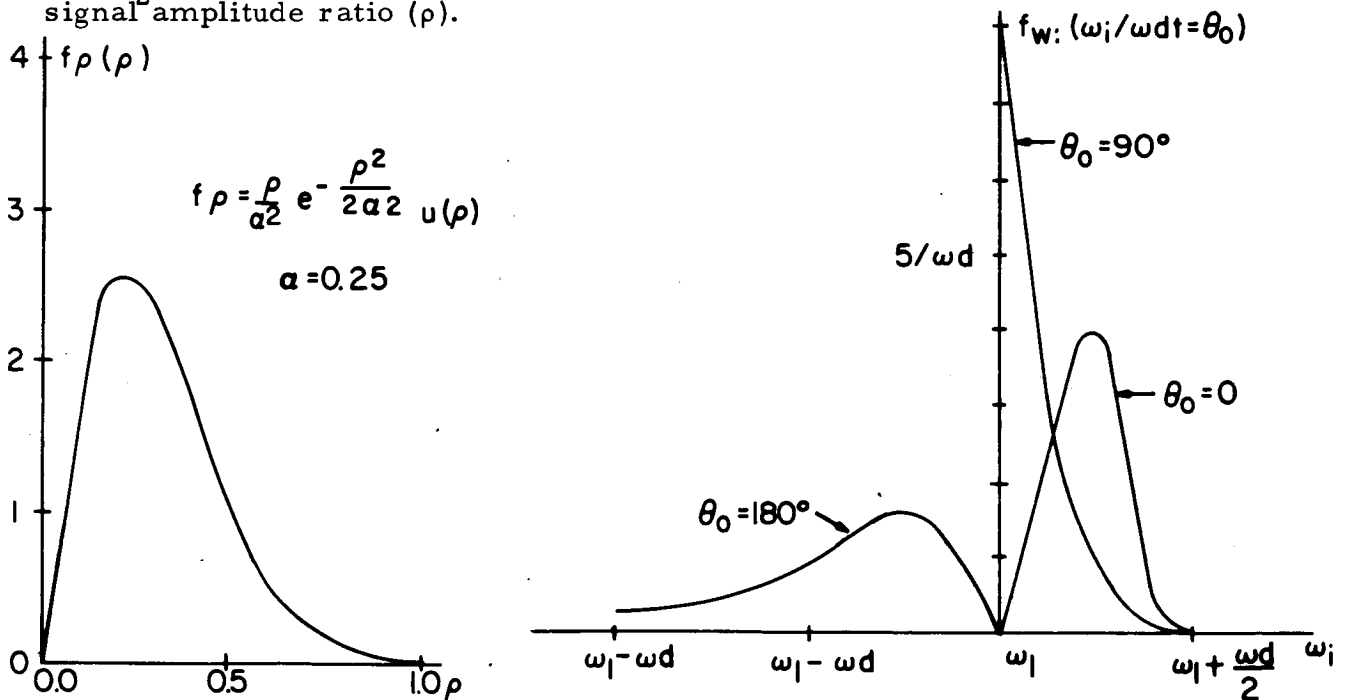
$$(\frac{\Delta\omega}{2})^2 \sin^2 \rho \frac{\Delta t}{2} \frac{\rho^2}{1-\rho^2} \Delta\omega = \rho\beta$$

This is an estimate of the power of the interference resulting from a single echo of an FM signal with monochromatic modulation. This is not the exact power resulting from a particular case, but rather an average over the waveform ensemble. It would be well to confirm the validity of the averaging by calculating the variance of this power over the ensemble, but the computation is not simple. At present, the best confirmation is that the waveforms do not vary drastically as the average phase difference, θ , varies (cf fig. 2(c) and fig. 2(e)).

A statistical approach is probably the only way to gain meaningful theoretical results from FM interference problems. The algebraic complexity of the results derived for general cases tends to hide any significance of the results. In multipath transmission, analysis is practical if only one spurious echo occurs. Study of modulation more complex than the monochromatic case is under way. Analysis of more than one echo has so far been unsuccessful.



(a) Interference resulting when two unmodulated carriers of frequency ω_1 and ω_2 are received. The waveform is plotted for various values of the signal amplitude ratio (ρ).



(b) A possible distribution (Rayleigh) of signal amplitude ratio.

(c) Corresponding density function of the interference (fig. (a), above at three instants in the interference cycle.

Fig. 1. Interference resulting from simultaneous reception of two unmodulated carriers, and analysis when the ratio of the two signals is a random number.

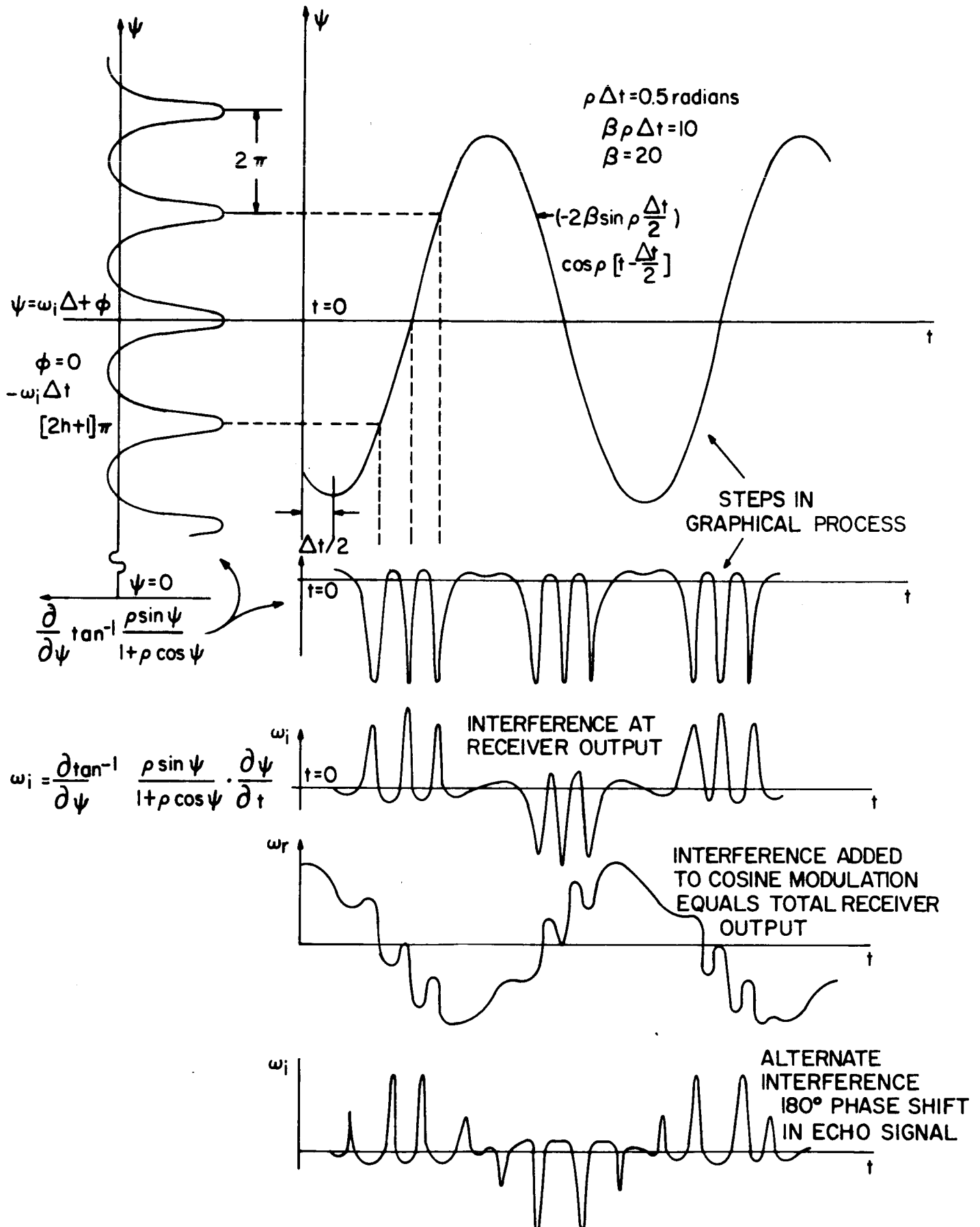


Fig. 2 Monochromatic modulation, one spurious echo FM multipath.

VI New and Continued Research

1. Optimum Demodulation

Techniques have been developed by the Communications group at PIB which can be used to determine the threshold response of the Maximum Likelihood Estimator (MAP) and the Least Mean Square Estimator (LMSE).

An iteration procedure has already been successfully demonstrated in the case of the MLE. Methods to shorten the computer time required will be investigated. The use of most likely noise is applicable to any FM demodulator having spikes at the output. The MLE has been shown to have multiple solutions spaced approximately 2π apart. This represents a spike. A computer solution of the LMSE will be sought and the MLE and LMSE compared. Lawton⁽¹⁾ has attempted this comparison using a scalar modulating signal. (We have already solved the MLE using a vector modulating signal).

The computer solution to the MLE is being studied to determine how best to reduce this to design a prototype circuit. This circuit will be constructed and tested and its performance compared to the computer results.

2. Sub-Optimum Receivers

Phase Locked Loops

Both the "most likely" noise model and the gaussian shaped, "constant amplitude" click model for noise will be used to determine the response of higher order phase locked loops, of various order FMFB systems and of FLL systems. Comparisons will be made both between systems and between types of loop filters within systems. Results will be obtained both for the case of no modulation and for the modulated case.

PLL Investigation Using Fokker-Plank Methods

The transient probabilities behavior of the phase locked loop was solved numerically using Fokker-Plank Methods. This is the only rigorous, non-approximate technique available and the results will be used to calculate cycle skipings which in turn can be related to the "clicks" mentioned throughout this report. An attempt will be made to extend the work to second order loops and should provide insight and some justification for the various approximate approaches.

Frequency Locked Loops

Extensive theoretical and experimental work is underway. The theoretical work involves calculations of the noise and distortion terms expected at the output of the FLL with either modulated or unmodulated noisy FM signals at the input. This work will allow theoretical optimum settings for loop bandwidths as well as providing theoretical data as to the effect of different types of loop filter. Independent experimental verification of the different portions of the output noise computed from the theoretical model will be undertaken.

In addition to the analog $a(t)$ channel of the original FLL a loop has been constructed with a quantized $a(t)$ channel. This version appears particularly well suited for combatting "spikes" caused by the transmission of FM through a fading channel. This is true since it is an experimental observation, made with the help of the water tank channel simulator, that such "spikes" are inevitably associated with every value for $a(t)$. Data will also be taken for output signal to noise ratios for ordinary analog signals and for error rates of FSK signals transmitted via both ordinary and quantized frequency locked loops.

Since it is shown in this report that the FLL provides a limiting case for the FMFB receiver, work will be continued to tie these devices together and to extend the FLL results to provide methods for designing FMFB receivers and in predicting their outputs.

Spike Detection

In this report we have shown that the spikes present at the output of a PLL can be detected. Circuitry is currently being developed to cancel these spikes and hence reduce the PLL threshold. Digital computer studies have demonstrated that spike cancellation is feasible.⁽²⁾

Spike detection has been applied to the reduction of errors in FSK.⁽³⁾ The improved procedures developed at PIB operating in conjunction with a PLL will improve results dramatically. Circuitry is currently being designed to demonstrate this.

3. Threshold in PCM

Some work has begun on the investigation of threshold effects in PCM. This work will continue and comparisons will be made of the performance of PCM, FM/FM and FM/PCM systems. Potential improvements possible in a PCM system by use of a feedback channel will also be investigated.

4. Synchronization Problems

As satellites move further from the earth, signals emanating from them become weaker. To maintain a low probability of error when transmitting data, the bits are transmitted at a very low bit rate. When this occurs the phase variation of the input signal must be considered. A program begun by the Communications group, to determine the probability of error under these conditions, to design synchronization systems to function in an optimal fashion, and to compare these optimal systems to those available today, is proceeding satisfactorily.

5. Proposed Work on the PIB Channel Simulator

The installation of a new PDP8 4000 word memory, digital computer with digital-to-analog and analog-to-digital capabilities will allow easy access to certain channel statistics that have not been obtained previously.

Work will continue with FM transmission of both analog and digital signals through the fading channel. Analog signals will include video from the PIB slow scan TV setup, multichannel test signals, and single channel test signals. Digital signals will include both ordinary FSK and various coded versions of the analog signals. The PDP8 will be used to perform certain of the coding, decoding, and error determination functions.

The work of comparing the performance of the various types of threshold extending FM receivers when receiving fading signals will continue.

Slow Scan TV System

Work shall continue utilizing the PIB slow scan TV system as a test signal to compare various threshold extending systems corrupted by fading and gaussian noise. In particular, a conventional limiter-discriminator, a Phase Locked Loop, Frequency Locked Loop, and a Frequency Demodulator with Feedback, are presented with the identical noise corrupted carrier which is modulated in frequency by the TV test signal and subjected to fading in the PIB fading channel. For each system the resultant output video display shall be obtained and compared with the others. This shall be accomplished for various carrier to noise ratios as well as various levels of fading.

6. Recursive Detection and Signal Design

The work already performed during the last year on recursive detection opens a vast territory for further research. The receiver approach can be applied to problems in detection as estimation theory including synchronous and non-synchronous detection, demodulation, diversity problems and many other important theoretical as well as practical problems of interest.

Because of the fact that explicit solutions may be obtained for quantities like signal to noise ratio and probability of error for any colored noise, it is possible to use the results for signal selection. Furthermore, since the problems are set up in state variable form, the form of the optimum detectors and demodulators are realizable in a relatively simple real time operation. This is especially important for digital implementation. Work will continue in the next year along these lines.

Both the "most likely" noise model and the gaussian shaped, "constant amplitude" click model for noise will be used to determine the response of higher order phase locked loops, of various order FMFB systems and of FLL systems. Comparisons will be made both between systems and between types of loop filters within systems. Results will be obtained both for the case of no modulation and for the modulated case.

VII

Masters Theses & Doctoral Dissertations

Masters Theses

The following Masters Theses of June 1967 graduates were partially supported by this grant:

1. Farrow, M. "Wideband FM Limiters"
2. Galbraith, W. "Computer Simulation of A Digital PLL"
3. Glassman, C. "S/N Ratio Through A Real Limiter"
4. Goldstein, A. "Arms Test Detector"
5. Guglielmetti, P. "A Median Detector"
6. Jankoniski, A. "Generations of Pseudo-Random Sequence of Symbols"
7. Kleinberg, L. "Likelihood Ratios for Deterministic Signals"
8. Kosovych, O. "Signal Design"
9. Kovacs, E. "Properties & Applications of Pseudo Random Sequences"
10. Kuhar, J. "Analysis of Smooth Bandpass Limiters"
11. Lemp, J. "A Digital Correlator"
12. Lutchinsky, J. "Pattern Recognizer"
13. Marko, J. "Wideband Sinusoidal FM Generator"
14. Maskasky, J. "FM System for Multichannel Transmission"
15. McDaniel, L. "Non-Linear Comparator for PLL"
16. Micheletti, T. "Ratio Detector"
17. Mishory, M. "Quadrature Modulations in Data Transmissions"
18. Moraitimis, N. "Analysis & Design of Transistorized AM Modulator"
19. Ribbens, J. "PSK Detection Employing PLL"
20. Unkauf, M. "Channel Simulation"
21. Yang, M. "FM Through Tuned Circuits"
22. Yavuz, D. "FM Spectra"

Ph.D. Theses

The following PhD Dissertations, of June 1967 graduates, were partially supported by this grant:

1. Crepeau, P. "Random Channel"
2. Kreussling, F. "Performance of M-ary Feedback Communication Systems"
3. Nelson, E. "Response of an FM Discriminator to an FM Signal in a Randomly Fading Channel"

VIII

Papers Published

The following papers were partially supported under this grant.

1. D. L. Schilling, K.K. Clarke, E.A. Nelson
"Analysis of an FM Discriminator with Fading Signal Plus Additive Gaussian Noise," Transactions on ComTech, April, 1967.
2. D.L. Schilling, E. Hoffman
"Demodulation of Digital Signals Using an FM Discriminator," NEC Conference, October 1966.
3. D.L. Schilling, M. Smirlock
"Intermodulation Distortion of a Phase Locked Loop Demodulator," Fourth Canadian Symposium on Communications, October 1966, and IEEE Trans. on ComTech, April 1967.
4. D.L. Schilling, F. Kreussling
"Application of Feedback to M-ary PAM," NEC Conference, October 1966.
5. D.L. Schilling
"Error Rates for Digital Signals Demodulated by an FM Discriminator," IEEE Transactions on Communication Technology, August 1967.
6. D.L. Schilling, A. Guida
"Extending the Threshold of the Phase Locked Loop," IEE Conference on Frequency Generation & Control, London, England, May 22-24, 1967.
7. D. L. Schilling, A. Guida
"The Optimal FM Discriminator Type Demodulator," 1967 IEEE International Conference on Communication, Minneapolis, Minnesota, June 12-14, 1967.
8. D.L. Schilling, A. Nelson
"The Response of an FM Discriminator to a Digital FM Signal in Randomly Fading Channels," IEEE International Symposium on Information Theory, San Remo, Italy, September 11-15, 1967. To be published in IEEE Trans. on ComTech.
9. D. L. Schilling
"The Detection of PCM/FM," - invited paper presented at the international Telemetry Conference, October 2-4, 1967.

Papers Published - continued

10. D.L. Schilling, J. Kuhar
"Signal to Noise Ratio for Bandpass Limiters," IEEE Transactions AES, February 1968.
11. K.K. Clarke, D.T. Hess
"Frequency Locked Loop FM Demodulator," IEE Transactions on ComTech, August 1967.
12. K.K. Clarke
"Fading Channel Simulators," Proc. IEEE Correspondence, January 1967.
13. R.L. Pickholtz, R.R. Boorstyn
"A Recursive Approach to Signal Detection," IEEE International Symposium on Information Theory, San Remo, Italy, September 11-15, 1967. To be published in IEEE Trans. on Info. Theory.